SEROTONIN NEURONS MODULATE LEARNING RATE THROUGH UNCERTAINTY

by
Cooper Donald Grossman

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
February, 2021

Abstract

Regulating how fast to learn is critical for flexible behavior. Learning about the consequences of actions should be slow in stable environments, but accelerate when that environment changes. Recognizing stability and detecting change is difficult in environments with noisy relationships between actions and outcomes. Under these conditions, theories propose that uncertainty can be used to modulate learning rates in a process known as meta-learning. To test these theories, we developed a mouse model of dynamic foraging in which probabilistic relationships between actions and outcomes change over time. We show that mice behaving in this task exhibit choice behavior that varied as a function of two forms of uncertainty estimated from a meta-learning model. In this model, *expected* uncertainty follows recent variability in action-outcome relationships and mitigates learning when those relationships are stable. *Unexpected* uncertainty monitors deviations from this expected variability in order to enhance learning when a change in the action-outcome relationship may have occurred. The activity of dorsal raphe serotonin neurons tracked both types of uncertainty in the foraging task, as well as in a dynamic Pavlovian task. The activity tracked uncertainty on both fast and slow timescales. Reversible inhibition of serotonin neurons in the foraging task reproduced changes in learning predicted by a simulated lesion of uncertainty-driven meta-learning in the model. The rate at which representations of action value in the medial prefrontal cortex were updated (learning rate) was also modulated by activation of local serotonin axons. We thus provide a quantitative link between serotonin neuron activity, learning, and decision making.

Thesis advisor and primary reader: Jeremiah Y. Cohen, Ph.D.
Second reader: Vikram Chib, Ph.D.

Dedicated to my family,

who have enabled and inspired me,

through profound love, support, and example,

to pursue what most interests me.

Acknowledgments

I am in debt to the community of wonderful people around me, in the Ph.D. program and beyond, for being able to complete this thesis and remain happy in doing so.

I am thankful for those who helped me get to the Ph.D. program to begin with. My first experience in neuroscience research was in the lab of Patricia Janak when her lab was at UCSF. I was incredibly fortunate, as a philosophy major with no neuroscience research experience, to be given the opportunity to work in an environment with such exceptional research and nurturing training. The knowledge I gained from working with Tricia, members of her lab, as well as members of neighboring labs has been invaluable. I am especially appreciative of the keen, patient mentorship and training I received from Tricia, Zayra Millan, Jeff Simms, and Woody Hopf. It was in this lab that I absorbed a deep appreciation for animal behavior and what it can tell us about the brain. When Tricia moved her lab to Johns Hopkins, I decided to stay in San Francisco (a decision that ended up being short-lived). Thanks to a fortunate interaction with Tosha Patel in the communal kitchen and an open technician job, I ended up joining the lab of Vikaas Sohal. I want to acknowledge Vikaas and Tosha for their encouraging support, insight, and training, Margaret Cuniff for teaching me *in vitro* whole-cell recordings and trusting me with her rig, Jiggy Athilingam also for teaching me how to patch and keeping me entertained during long hours at the rig, and Anthony Lee as well as Jillian Iafrati for mentorship on experimental design and very enjoyable bike rides.

Despite the best efforts of these admirable scientists and through no fault of theirs, I entered into this Ph.D., like many students, with pretty severe imposter syndrome. Part of overcoming this mental state involved developing an expertise that allowed me to complete the

work presented in this thesis. The other part was continuing to develop an understanding of the humanity of scientific research and a confidence in my own abilities. These developments are the product of the scientists, friends, and loved ones in my life.

First, I want to thank my mentor, Jeremiah Cohen. Jeremiah has a special talent for regarding his students and post-docs with trust and esteem in a way that fosters confidence. After almost-weekly meetings with Jeremiah, I would come away with my anxieties assuaged in addition to having a clearer sense of goals and reinvigorated motivation to do my work. His excitability about science is far-reaching (extending well beyond science) and contagious. This breadth in curiosity underlies his enlightened, comprehensive perspective on science and how to conduct it. I also appreciate Jeremiah's genuine recognition and care for his lab members as humans outside of their roles as scientists. He has a refreshing and understanding perspective on how academia should progress on such issues, and is an example of it. I have learned a great deal about rigorous, thoughtful science and compassionate mentorship from Jeremiah.

My friends and labmates, Bilal Bari, Fede Lucantonio, Eunyoung Kim, Anna Chang, Sue Su, Helia Seifikar, Matt Lewis, and Ahmad Taha are also to thank. These lovely humans made going to lab even more rewarding and edifying. I want to acknowledge Bilal Bari, especially, for his substantial contributions to the development of this project, my amusement, and my sanity. I also want to thank the friends and collaborators in the neighboring labs of Dan O'Connor and Gül Dölen. Eastman Lewis and Romain Nardou were a pleasure to work and eat lunch with during my rotation in the Dölen lab and well after.

Mentorship from my thesis committee has also been critical to the development of this thesis and my growth as a scientist. My first meeting with Tricia, Solange Brown, and Peter Holland was intimidating but soon into the meeting I was reassured by their encouragement and the comforting way in which feedback and critique were delivered. After Peter retired, Vik Chib joined the committee and fit right into this approach. I am grateful also for their active engagement, taking the time to understand and improve my research. The insights

and unique perspectives that each member brought to this meeting have shaped this project and my approach to science.

One of the main reasons I chose to join the neuroscience program at Johns Hopkins was the fun, welcoming, talented, intelligent, down-to-earth, and genuine members of the student body. My time in the program would not have been nearly as enjoyable or enlightening without them. Gabby Sell, my big sibling and softball manager, made sure I was welcomed and informed from the first day and throughout. My friends and roommates David Ottenheimer and Lionel Rodriguez were constant sources of amusement and insightful conversation. The Whisk(e)y Club provided an excuse for Akash Khanna, Bryce Grier, Cody Call (among others), and I to spend delightful and life-enriching time together. Good times and much laughter were also made possible by Raina D'Aleo, Althea Cavanaugh, Alice Berners-Lee, Dallas Khamiss, and too many other friends to list.

The last two years of my Ph.D. were brightened by my partner, Jing Liu, who I want to thank. Your encouragement and support have helped me to complete this work with joy. The most productive periods of writing of this thesis were done in your company, thanks to your reassuring presence and motivation to spend more time with you. You have made my time outside of lab fulfilling with your care, generosity, and stress-melting goofiness.

And finally, I want to express my gratitude for my family. Your love and support has been so foundational for me that it is impossible to imagine a life without it, much less a Ph.D. Your boundless encouragement and enabling of my goals has allowed me to pursue my curiosities to the utmost. And the examples you all have set have given me something to aspire to. Throughout graduate school, regular phone calls with each of you always brought a stabilizing peace and reassurance that made me feel at home (from across the country) and reoriented my perspective to the most important things. Thank you all, for everything.

Contents

List of Figures

Chapter 1

Introduction

Serotonin is a simple indolamine molecule that has, for billions of years, been created by organisms for various biological purposes. While its presence and function in early single cell organisms is hard to discern from its close metabolite tryptophan, serotonin is demonstrably present in almost all living plants and animals, as well as in some fungi and single-celled eukaryotes (Wier and Tyler, 1963; Garattini and L., 1965; Saxena et al., 1966; Smith, 1971; Azmitia, 1999). Given its ancient roots and existence within vastly different biological systems, it is of no surprise that its associated functions are highly varied. Even just within the human body serotonin pervades every organ and is thought to be involved in the regulation of gastrointestinal function (Vialli and Erspamer, 1937; Spencer et al., 2015; Keating and Spencer, 2019; Jones et al., 2020), hematopoietic function (Baumgartner and Born, 1968; Walther et al., 2003; Mammadova-Bach et al., 2018) and other roles in cardiovascular system (Rapport et al., 1948; Vanhoutte, 1987), and periperheral metabolism (Young et al., 2015; Martin et al., 2017). Of its many roles, the most complex may be its function in the human central nervous system. Dense axonal arborizations from serotonin neurons innervate almost the entirety of the brain and spinal cord. The activity of the neural serotonin system has been associated with regulating physiological state, like thermoregulation (Feldberg and Myers, 1963; Morrison and Nakamura, 2011; Ishiwata et al., 2017), arousal and sleep-wake cycles (McGinty and Harper, 1976; Trulson and Jacobs, 1979; Yuan et al., 2005; Monti, 2011), feeding, and drinking (Blundell, 1977; Ribeiro-do Valle et al., 1989; Simansky, 1996; Lee and

Clifton, 2020). The system also appears to be involved in cognitive functions like learning and decision making (Soubrié, 1986; Winstanley et al., 2004; Clarke et al., 2004, 2007; Cools et al., 2008; Bari et al., 2010; Matias et al., 2017), sensory perception (Ranade and Mainen, 2009; Dugué et al., 2014; Lottem et al., 2016; Seillier et al., 2017), pain (Palazzo et al., 2004, 2006; Neugebauer, 2020), mood (Harmer et al., 2004; Andrews et al., 2015; Godlewska et al., 2016; Michely et al., 2020), and social interaction (Wood et al., 2006; Crockett et al., 2008; Dölen et al., 2013; Lee and Goto, 2018).

While this heterogeneity makes understanding the precise function of serotonin neuron function a daunting task, some have proposed theories of a unified function for serotonin in the central nervous system. One of these theories suggests that serotonin neurons serve to regulate behavioral state, predisposing the animal to certain sets of behaviors based on internal state (e.g., level of thirst) and environmental conditions (e.g., predatory threat level; Jacobs and Fornal, 1991). Others have referred to the ability to adapt behavior to changes in environmental conditions as behavioral flexibility (Clarke et al., 2004, 2007; Matias et al., 2017). Similarly, some have called these processes a maintenance of homeostasis and have linked this integration of brain, body, and world to homeostatic mechanisms at the synaptic level (Azmitia, 2001, 2007). Extensive literature supports these general notions but the field is far from a comprehensive, mechanistic description of serotonin neuron function—one that quantifies what information is received by these neurons, how it is computed, and how it is propagated downstream to modulate brain function at various levels of neural organization, cognition, and behavior. Since dysfunction of the serotonin system has been implicated in the pathology of schizophrenia, mood, anxiety, eating, and addiction disorders, having such a thorough understanding of serotonin neuron function is crucial to the development of better therapies for these psychiatric disorders. The most effective pharmacological treatment for major depression, for example, is selective serotonin reuptake inhibitors which increase extrasynaptic levels of serotonin without spatial or temporal specificity. Despite their relative efficacy, amelioration of depressive symptoms takes weeks to actualize, a time during which

those symptoms and risk of suicide actually increase. Without a more extensive knowledge of serotonin neuron function, in health and disorder, more precise and effective treatments cannot be designed.

In developing a more precise understanding of the serotonin system, research has increasingly focused on certain subpopulations of serotonin neurons. This narrowing of focus is not only a natural consequence of the reductive nature of the scientific enterprise but is motivated also by the striking diversity of these neurons. The majority of serotonin neurons are found in 9 raphe nuclei in the midbrain and brainstem (Dahlström and Fuxe, 1964) and demonstrate an assortment of genetic expression profiles, anatomies, and electrophysiological features (Steinbusch, 1981; Ishimura et al., 1988; Aitken and Törk, 1988; Baker et al., 1990; Törk and Hornung, 1990; Törk, 1990; Baker et al., 1991; Jacobs and Azmitia, 1992; Jensen et al., 2008; Alonso et al., 2013; Okaty et al., 2015, 2019, 2020). Much attention has been paid to the dorsal raphe which, along with the median raphe, innervates the majority of the forebrain and provides most of the serotonin in the brain. This structure is also amenable to study in model organisms since it is relatively homologous across mammals, including humans (Baker et al., 1990; Hornung, 2003). Among many functions, studies of dorsal raphe serotonin neurons have associated them with motivated behavior. The results from these studies suggest a specific role in learning about the relationships between stimuli or actions and valued outcomes. The generality of such a cognitive function may provide a unifying explanation of some of the various behaviors with which serotonin has been associated. Even if seeking a unifying function is a fool's errand, motivated behavior provides one avenue for understanding one facet of serotonin neuron function. Most of these previous results, however, come from manipulations of serotonin neurons that are temporally and spatially imprecise. Additionally, very few studies have observed the activity of identified dorsal raphe serotonin neurons during learning. As such, a mechanistic understanding of serotonin neurons in learning remains to be determined.

Elucidating the mechanisms of learning is difficult, in part, because learning is a cognitive

process that operates on latent variables not directly observable in behavior. The brain can be thought of as a black box in this way, taking in information about the environment and transforming it into observable behavior in order to achieve its internal goals. We can seek to understand the hidden process by devising quantitative models that propose a mathematical explanation of that transformation. This approach has been leveraged to provide insight into the function of dopamine neurons (Schultz et al., 1997; Cohen et al., 2012), basal ganglia (Samejima et al., 2005; Lau and Glimcher, 2008; Cai et al., 2011; Wang et al., 2013), and neocortex (Tsutsui et al., 2016; Bari et al., 2019). However, few models have been proposed to specifically explain the role of serotonin in learning (Doya, 2002; Daw et al., 2002) and the predictions of these models have not been thoroughly tested.

The body of work presented in this thesis seeks to provide a quantitative link between dorsal raphe serotonin neuron function, learning, and decision making behavior. First, in this introduction I will review the anatomy of dorsal raphe serotonin neurons and serotonin receptors. Then, I will review the various associated functions, elaborating on the neuromodulatory system's specific role in the context of learning about rewards. I will also introduce basic computational models of learning and decision making and how they can be modified to explain observed behavior as well as serotonin neuron function. And given the promiscuous nature of serotonin neuron axons in the brain, I will also introduce the functional role of the prefrontal cortex—a region of the brain involved in learning and decision making—which may implement serotonin neurons' effects on learning. In subsequent chapters, I will describe experiments examining serotonin neuron activity, in the dorsal raphe and prefrontal cortex, in the context of learning and decision making. I will also define and characterize the computational models we designed to explain this activity and behavior. Finally, I will discuss the various interpretations of these results, their limitations, and their significance in the context of serotonin neuron research.

1.1 Dorsal raphe serotonin neuron and serotonin receptor subtypes and anatomy

Serotonin was first discovered in mammalian biology in 1937 through its extraction from enterochromaffin cells and observed effects on smooth muscle contraction in the intestines (Vialli and Erspamer, 1937). In the 1960's, novel histochemical techniques allowed for the first demonstration of serotonin-containing cell bodies in the midbrain and brainstem (Dahlström and Fuxe, 1964). Since that landmark study, the anatomy of serotonin neurons and their receptors have been studied extensively across species. While the anatomical delineations have evolved over the course of this research, it is generally agreed upon that serotonin neurons in the brain are spread across 9 raphe nuclei (B1-B9), defined by their developmental origins, projection targets, and genetic identities (Steinbusch, 1981; Ishimura et al., 1988; Aitken and Törk, 1988; Baker et al., 1990; Törk and Hornung, 1990; Törk, 1990; Baker et al., 1991; Jacobs and Azmitia, 1992; Jensen et al., 2008; Alonso et al., 2013; Okaty et al., 2015, 2019).

The dorsal raphe comprises 2 of the raphe nuclei, consisting of neurons rostral (B7) and caudal (B6) to the isthmus (Dahlström and Fuxe, 1964; Ishimura et al., 1988; Baker et al., 1990, 1991; Jacobs and Azmitia, 1992). The dorsal raphe contains the largest number of serotonin neurons of the raphe (165,000 in human dorsal raphe (Baker et al., 1991) and 9,000 in mouse dorsal raphe (Ishimura et al., 1988)), but even in total serotonin neurons represent less than 0.1% of all neurons in the brain (Halliday et al., 1988; Baker et al., 1990, 1991; Ishimura et al., 1988; Jacobs and Azmitia, 1992). One of the reasons such a small group of neurons has received such scrutiny is due to their pervasive innervation of the vast majority of the central nervous system (Figure 1-1). Serotonin neurons in the dorsal raphe alone send axonal projections to most of the forebrain structures (Jacobs and Azmitia, 1992). These regions include thalamus (Vertes et al., 2010), hypothalamus (Chowdhury and Yamanaka, 2016), basal forebrain, amygdala (Linley et al., 2017; Ren et al., 2018),

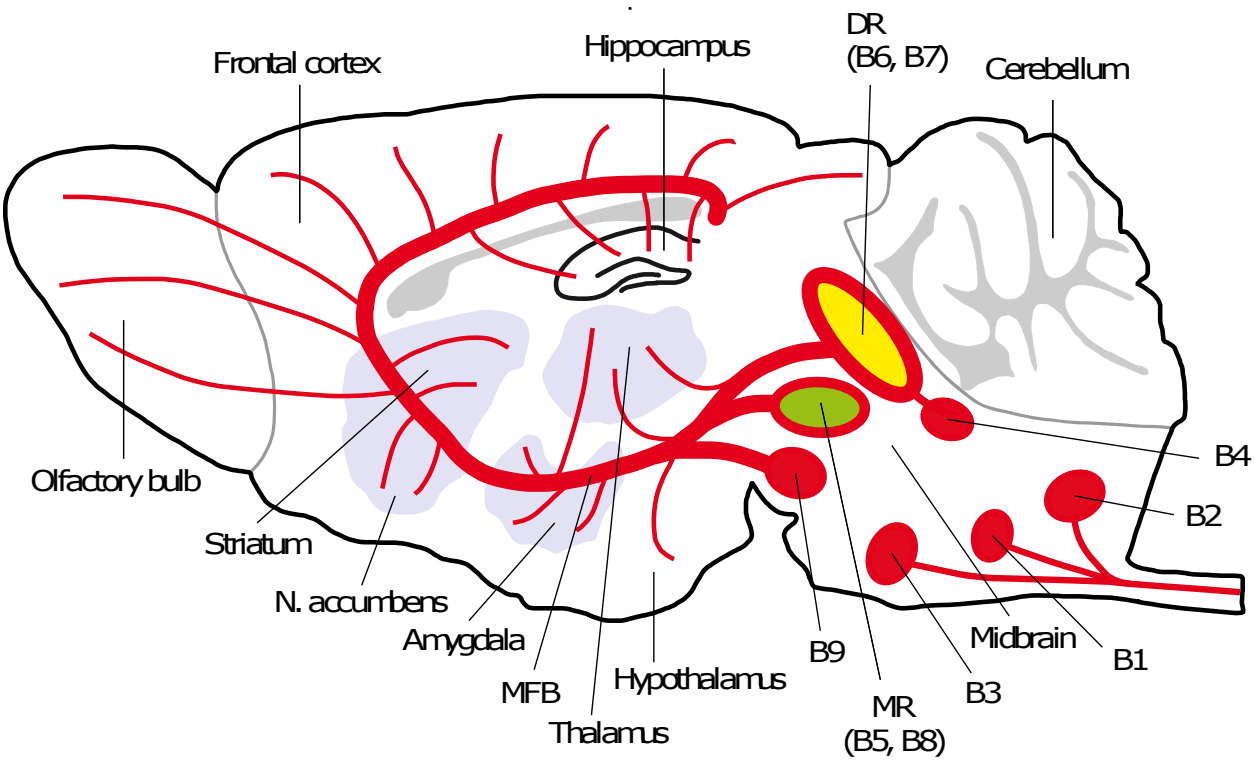

Figure 1-1: **Rodent brain serotonin system. Figure and legend adapted from Lesch and Waider 2012.** Serotonin neuron clusters are organized in the 9 raphe nuclei, B1–B9. The more caudal nuclei (B1–B3) in the medulla project axons to the spinal cord and the periphery, whereas the more rostral raphe nuclei contain the principal dorsal raphe group (B6 and B7; depicted in yellow) and the median raphe group (B5 and B8; depicted in green), which project to different but overlapping brain areas. DR, dorsal raphe nucleus; MFB, medial frontal bundle; MR, median raphe nucleus.

nucleus accumbens (Brown and Molliver, 2000), dorsal striatum, ventral tegmental area, and throughout cortex (Linley et al., 2013; Prouty et al., 2017; Ren et al., 2018) among other subcortical structures. These projection patterns are largely distinct from targets of other raphe nuclei, like the median raphe which also innervates the forebrain (Vertes, 1991; Vertes et al., 1999; Vertes and Linley, 2007, 2008). Serotonin neurons in the dorsal raphe show some topography based on their projection targets (Bang et al., 2012; Muzerelle et al., 2014; Fernandez et al., 2016; Prouty et al., 2017). Retrograde viral tracing of subsets of these neurons confirmed some topography, but also revealed that these innervation patterns are complex (Ren et al., 2018). Neurons that innervate the orbitofrontal cortex have a ventral bias in dorsal raphe but also send collaterals to olfactory bulb, entorhinal cortex, and cortical

amygdala. Central-amygdala-projecting neurons were mostly found in the dorsal part of the dorsal raphe. Reconstruction of single neurons expanded upon this observation, showing more target specificity of subcortically-projecting neurons, but substantial collateralizations in cortex-projecting cells (Ren et al., 2019).

Inputs to dorsal raphe serotonin neurons have also been mapped and show less promiscuity than their outputs (Weissbourd et al., 2014; Pollak Dorocic et al., 2014; Ogawa et al., 2014). Using rabies viral techniques to determine monosypatic inputs to genetically identified serotonin neurons, substantial inputs were found to come from different regions in the prefrontal cortex, extended amygdala, lateral habenula, hypothalamus, substantia nigra, ventral tegmental area, and the medulla (Weissbourd et al., 2014). Serotonin neurons also receive local input from other serotonin neurons and GABA neurons. Some studies have demonstrated the circuit-level function of these inputs. Prefrontal cortex inputs to dorsal raphe activate serotonin neurons directly followed by feed-forward inhibition through activation of local GABA neurons (Geddes et al., 2016; Pollak Dorocic et al., 2014; Zhou et al., 2017). Lateral habenula inputs were also shown to drive excitatory responses in serotonin neurons (Pollak Dorocic et al., 2014). Inputs from D1 dopamine receptor expressing neurons in the caudoputamen drove GABA-mediated inhibitory responses (Pollak Dorocic et al., 2014).

Differences in anatomy across individual neurons suggests the possibility of distinct subtypes of serotonin neurons. This idea is supported by considerable variability in their genetic profiles (Huang et al., 2019; Ren et al., 2019; Okaty et al., 2020). Genetic fate mapping studies demonstrate that they have various developmental origins across several different rhombomeres—the transiently discrete segments of the neural tube in the developing hindbrain (Jensen et al., 2008; Alonso et al., 2013; Niederkofler et al., 2016; Okaty et al., 2015; Ren et al., 2019; Okaty et al., 2020). Differences in genetic expression also reveal distinct molecular identities, even within dorsal raphe serotonin neurons from the same rhombomeric domain (Jensen et al., 2008; Alonso et al., 2013; Niederkofler et al., 2016; Okaty et al., 2015;

Ren et al., 2019; Okaty et al., 2020). Clustering analyses in some of these studies have suggested as many as 14 molecularly-defined subtypes of dorsal raphe serotonin neurons (Okaty et al., 2020), although the patterns of expression are overlapping and some of these distinctions may be partially a result of methodology rather than biology. Some of the clusters are clearly distinct, such as those neurons that are enriched for the transcript *P2ry1*, which encodes for purinergic receptor P2Y1. The chemical identity of these neurons maps onto a specific topography as well, with these neurons residing in the dorsal, caudal portion of dorsal raphe. Other genes are expressed across multiple clusters, but also have spatial biases in the dorsal raphe. The *Vglut3* (vesicular glutamate transporter gene) enriched neurons, for example, makeup about %60 of serotonin neurons and are generally found in the ventral portion of dorsal raphe. While the spatial distributions of dorsal raphe serotonin neurons with different genetic expression profiles are not as clearly delineated as in cortical layers, the substantial organization does point to classifiable subtypes. This idea is further supported by the same studies, among others, that have shown that the topography in the dorsal raphe also matches efferent projection targets. The *P2ry1* enriched neurons are supra-ependymal projecting, while the *Vglut3* expressing neurons predominantly target the olfactory bulb and cortex(Ren et al., 2019; Okaty et al., 2020).

These genetically- and anatomically-diverse neurons are unified by their ability to make and release serotonin. Serotonin is released at defined synapses via axons, but also extrasynaptically along axons, dendrites, and from the cell bodies (Descarries et al., 1975; De-Miguel et al., 2015; Parent and Descarries, 2020). In addition to multiple release sites, post-release responses are determined by at least 14 different serotonin receptor subtypes (in addition to splice and gene-edited variants) belonging to 7 different families (Vilaró et al., 2020; Barnes et al., 2021). Except for the ionotropic 5-HT_3 subtype, serotonin receptors are G-protein-coupled receptors that mediate both excitatory and inhibitory effects on excitability, synaptic plasticity, and other postsynaptic consequences (Derkach et al., 1989; Barnes et al., 2021). The 5-HT_{1A} receptor, for example, is expressed postsynaptically and

hyperpolarizes the neuron by enhancing inhibitory currents through gated inwardly rectifying potassium channels, among other effects (Andrade and Nicoll, 1987). Another species of the same receptor is expressed somato-dendritically on serotonin neurons themselves and arbitrates inhibitory effects on excitability through the same mechanism (Aghajanian and Lakoski, 1984; Montalbano et al., 2015). Binding to the 5-HT_{2A} receptor, on the other hand, induces persistent enhancements of excitability (McCormick and Wang, 1991; Sheldon and Aghajanian, 1991). Activation of this receptor triggers several intracellular, Gα_q-dependent and independent signaling pathways and has been shown to modulate synaptic plasticity by phosphorylation of presynaptic NMDA receptors and postsynaptic AMPA receptors (Barre et al., 2016; Berthoux et al., 2019). 5-HT_{2C} receptors have also been shown to modulate plasticity through effects on receptor desensitization and internalization (Bécamel et al., 2004).

The relative diversity of serotonin receptor subtypes has allowed for nuanced specialization of the serotonin signal across cell types, circuits, brain regions, and networks. 5-HT_{3A} receptor subunit mRNA, for example, is found somewhat selectively in certain regions that include hippocampus, amygdaloid complex, septal region, olfactory regions, and neocortex (Tecott et al., 1993). In these areas in particular, 5-HT_{3A} receptors are found in GABAergic neurons (Morales and Bloom, 1997). In medial prefrontal cortex, these neurons inhibit other inhibitory GABAergic neurons, resulting in disinhibition of glutamatergic pyramidal neurons (Santana and Artigas, 2017; Dale et al., 2018). Other subtypes follow regional patterns of expression as well. Of relevance to this dissertation, cortical and limbic areas demonstrate comparatively high expression of 5-HT_{1A}, 5-$HT_{2A/C}$, 5-HT_3, 5-HT_4, and 5-HT_6 receptors (Barnes et al., 2021). The basal ganglia are enriched for 5-HT_{1B}, 5-HT_4, and 5-HT_6 receptors. In addition to being present in specific populations and regions, specialization by receptors is also achieved by expression of multiple receptor subtypes in the same neuron. Layer 5 neurons in the medial prefrontal cortex express both 5-HT_{1A} and 5-HT_{2A} receptors (Ashby et al., 1994). The relative expression of these receptors leads to excitatory, inhibitory, and biphasic responses

to serotonin (Araneda and Andrade, 1991; Puig et al., 2005; Avesar and Gulledge, 2012; Stephens et al., 2014).

Serotonin neuron and receptor diversity reveals the astounding complexity of the serotonin system. This summary betrays some of that complexity in its brevity and also ignores serotonin receptor expression in glia, presence of other neurotransmitters in serotonin neurons, interactions between serotonin receptors and receptors for other neurotransmitters, and other anatomical considerations. The findings summarized here are those most relevant to the research presented in this dissertation.

1.2 Serotonin neuron function

Unifying disparate serotonin neuron functions

Through various types of manipulations, the serotonin system has been implicated in a large number of behavioral functions. Yet, in all of these cases the effects of modulating serotonin are small and animal behavior is largely intact. For these reasons, B.L. Jacobs said that serotonin "it is at once implicated in virtually everything, but responsible for nothing" (Jacobs and Fornal, 1995). While evolution and natural selection are imperfect engineers, it seems unlikely that such a metabolically costly system of neurons—that pervades the vast majority of the central nervous system—would be unnecessarily preserved. In light of considerations about efficiency, it is also possible that this system has been repurposed for various functions over the course of evolution. The observation that serotonin neurons are not essential for anything may be partially explained by the interactions of the different neuromodulatory systems and their ability to compensate. This degeneracy may actually be indicative of their fundamental importance to healthy brain function (Edelman and Gally, 2001).

Regardless, these varied effects warrant understanding. One possibility is that each associated behavioral function is realized by a distinct subset of serotonin neurons, an idea that might be supported by the diversity of serotonin neurons in genetic expression profiles,

anatomy, and electrophysiological characteristics(Huang et al., 2019; Ren et al., 2019; Okaty et al., 2020). There is evidence for functional differences in subpopulations as well (Hale and Lowry, 2011). The indirect marker of cellular activity, c-Fos, was used to distinguish subsets of serotonin neurons with different projection targets that were preferentially activated by stress, for example (Otake et al., 2002; Hale et al., 2008). At the same time, this diversity may reflect contingencies at the level of biological implementation, while ultimately the neurons perform the same computational function. Theoretical work from Prinz and Marder, for example, demonstrates that models of simple crustacean stomatogastric ganglia robustly produce similar network activity across different combinations of cell properties and synaptic strengths (Prinz et al., 2004). Similarly, different serotonin neurons are known to encode the same information with opposite changes in firing rates, obscuring effects of gross manipulations of the population (Liu et al., 2014; Cohen et al., 2015; Li et al., 2016; Ren et al., 2018). These possibilities are also not mutually exclusive. Serotonin neuron function could be described by some general function, like regulation of behavioral state, while subsets of neurons play more specific roles.

A satisfying and useful explanation of serotonin function will have to be mechanistic, even if it describes a multitude of separate processes. In this pursuit, much of the early research (which found such far-reaching behavioral associations) was limited by the methodology of the time. Lesions of cell bodies or axons are temporally imprecise, allowing more time for compensatory mechanisms. Pharmacological activation or inactivation of specific receptor subtypes do not mimic endogenous release, especially in the case of systemic administration.

Dorsal raphe single neuron electrophysiology during behavior

While manipulations are important for demonstrating causality between activity and behavior, determining the computations of these neurons requires recording action potentials during behavior. Early electrophysiological recordings from dorsal raphe laid the groundwork for understanding the activity of serotonin neurons. The first of these studies recorded from cats

and found that dorsal raphe neurons were active during quiet wakefulness, diminished during slow wave sleep, and completely ceased firing during rapid-eye-movement sleep (McGinty and Harper, 1976; Trulson and Jacobs, 1979). Further, these neurons were increasingly active during active waking, in an aroused state, and showed phasic bursts in response to sudden auditory stimuli that elicited an orienting response (Trulson and Jacobs, 1979). Later studies showed that these phasic responses generalized to visual stimuli and did not habituate over repeated presentations of either type of stimulus (Heym et al., 1982; Rasmussen et al., 1986). In regards to behavior, early studies also showed a relationship between dorsal raphe neuron firing rates and oral-buccal movements like masticating, biting, drinking, and oral grooming (Ribeiro-do Valle et al., 1989). The researchers also found increases in firing rates just prior to these behaviors, for example, when cats could see or smell the food they were about to chew.

While not recognized at the time, these observations hinted at a relationship of dorsal raphe neuron activity to valued outcomes, like food and water. After a couple decades of pharmacological and lesion studies suggested a role for the serotonin in motivated behaviors, a few groups began observing responses of dorsal raphe neurons to valued outcomes and the cues that predict them. Nakamura et al. recorded extracellularly from dorsal raphe neurons in rhesus monkeys performing a two-alternative forced choice, rewarding saccade task (Nakamura et al., 2008). In this task, animals fixated on a central point before being instructed to make a saccade to the left or right by a visual stimulus. The direction of the saccade predicted either a large or small reward and these contingencies switched after blocks of trials. The activity of many dorsal raphe neurons was responsive to the rewards and the cues that preceded them. Some neurons increased their activity at the time of the fixation cue and continued to increase until the time of the reward. The activity increased further when a large reward was delivered and decreased when a small reward was delivered. Other neurons showed the inverse response profile with inhibitory responses. The changes in firing rates at the time of the fixation cue seemed to imply the representation of the expected value.

Interestingly, this was an expectation that generalized across both left and right actions and their associated outcomes, since the animal did not yet know which instructional cue would be presented. An expectation of this type then, would be the average of the large and small rewards and possibly signals state value. The activity of many neurons also tracked the expected value at the time of cue that instructed which saccade to make. Here, the activity reflected the high or low value predicted by the cue and adapted accordingly when the contingencies between saccade direction and reward size were switched. Activity at the time of the outcome was prominent and was scaled by the size of the reward. A follow-up study used a similar task design but delivered no rewards instead of small ones (Bromberg-Martin et al., 2010). This task modification demonstrated that dorsal raphe neurons responses to no rewards were similar to small rewards, further bolstering the idea that that neurons were responding to outcomes relative to an expectation, and not just the rewards *per se*.

Subsequent work from Nakamura's group continued testing the idea that dorsal raphe neurons' activity related to expected value and sought to examine divergences from it at the time of the outcome (Hayashi et al., 2015). In doing so, they implemented a Pavlovian version of the task in which the fixation cue was followed by one of three conditioned stimuli that predicted reward, reward at 0.5 probability, and no reward, respectively. The outcome was delivered at some delay without any dependence on monkey behavior. Many neurons responded to the conditioned stimuli with firing rate changes graded with the value of the predicted outcome. Again, they observed that separate groups of these neurons responded with opposite changes in firing rates, with some neurons preferring rewards and others preferring no reward outcomes. When outcomes following the 0.5 probability conditioned stimulus were delivered, these groups of neurons responded with opposite changes to the different outcomes that indicated value relative to the average, expected value. To be clear, this was not activity indicative of a reward prediction error, which is the *difference* between expected value and outcome that has been observed in dopamine neurons.

The same study also used blocks of the same task structure but with aversive puffs of air

to the face instead of rewards (Hayashi et al., 2015). In these trials, very few dorsal raphe neurons reacted to the conditioned stimuli, but many reliably responded to the airpuff itself. The response was rarely modulated by the probability of the airpuff. Interestingly, many neurons demonstrated long timescale, tonic changes in firing rates that differentiated between aversive and rewarding blocks. While the behavior of the monkeys differed between the types of blocks (licking or blinking), the observation is also consistent with the activity representing the value of behavioral context. Unlike dopamine neurons, dorsal raphe neuron activity signaled different aspects of value on distinct, but both behaviorally-relevant, timescales.

Observations of dorsal raphe neurons in a two-alternative forced choice task in rats were similar to those observed in monkeys (Ranade and Mainen, 2009). After poking into a central port, a conditioned odor stimulus instructed rats to poke in either the left or the right port for a reward that was delivered at various delays. Just as in monkeys, the observed responses were significantly heterogeneous, but large numbers of neurons were responsive to one or multiple features of the task or behavior. Some groups of neurons showed changes in firing rates (excitation or inhibition) as the rats entered the central odor port, ramping activity as mice moved to the reward ports, or phasic reactions to reward delivery. A small group of neurons also responded to rare reward omissions. These results provide support of the relationship between dorsal raphe neuron activity, value expectations, and experienced value relative to those expectations.

Rat dorsal raphe neurons were also recorded in task in which an auditory cue predicted the timing of an outcome, but a contextual cue (house light on or off) predicted whether the outcome was a reward or no reward in blocks of trials (Li et al., 2013). The majority of neurons responded phasically to cues, with more of this group responding selectively or with larger responses in the no reward blocks. Some neurons also demonstrated phasic response aligned to the first anticipatory lick or to each lick in the lick bout. Differences in tonic activity between the cue and outcome were also seen between blocks. These findings are largely consistent with those that came before, but show how value-related signals generalize

to Pavlovian cues and cues of a different sensory modality.

Dorsal raphe recordings revealed substantial diversity and task-relevant responses in the context of value processing and reward-driven behavior. Mirrored excitatory and inhibitory changes in firing rates provided the basis for skepticism about systemic manipulations of serotonin neuron activity. Specific conclusions about serotonin neurons were limited, however, by the inability to distinguish those neurons from their non-serotonergic neighbors.

Serotonin neuron electrophysiology during behavior

Originally, researchers identified electrophysiological criteria to distinguish serotonin neurons from non-serotonergic neurons using intracellular recordings that were carried out in animals under anesthesia. Recorded cells were labeled and identified *post-hoc* by intracellular infusion of markers through the micropipette used for recording and subsequent antibody staining for identifying proteins. For some time, researchers recording extracellularly used these physiological criteria to identify putative serotonin neurons. However, it was later apparent that these criteria were not accurate, being subject to both type I and type II error (Allers and Sharp, 2003; Kocsis et al., 2006).

Pharmacological techniques have also been used to identify serotonin neurons *in vivo*. A few studies used the 5-HT_{1A} agonist 8-OH-DPAT, which suppresses serotonin neuron firing rates (Aghajanian et al., 1978; Vandermaelen and Aghajanian, 1983; Miyazaki et al., 2011). In one of these studies, activity of serotonin neurons related to expectations of valued outcomes was recorded in rats waiting extended periods of time (2 - 20 s) for food or water (Miyazaki et al., 2011). Single neurons exhibited tonic changes in firing during the waiting period that showed only some decay over long intervals. The magnitude of this decay predicted whether or the not the rat successfully waited until reward delivery. On trials in which rewards were omitted, the activity steadily decayed prior to the rats leaving the reward port. There was also phasic activity at the time of the rewards. These findings further bolstered the idea that serotonin neuron activity was related to expectations about valued outcomes and indicated

the possible behavioral relevance of such a signal.

It was only recently that genetic tools allowed for high-confidence identification of serotonin neurons in extracellular recordings in mice. Using genetic mutants that express cre-recombinase under the control of a target promoter gene, Cre-dependent viruses (or crossing animals with Cre-dependent mutants) can be leveraged to express exogenous proteins in the target population. One group of proteins that can be expressed are opsins: light-sensitive proteins that include ion channels and G-protein coupled receptors. Channelrhodopsins are a family of these proteins that conduct cations in response to certain wavelengths of light (Nagel et al., 2003; Boyden et al., 2005). By expressing channelrhodopsin in a genetic population of neurons, one can identify extracellularly recorded neurons by assessing the response of those neurons to laser stimulation (Cardin et al., 2009; Lima et al., 2009; Cohen et al., 2012).

To date, only a few publications have used this optogenetic technique to record from identified serotonin neurons (Liu et al., 2014; Cohen et al., 2015; Li et al., 2016). Inspired by recordings from regions containing serotonin neurons, these studies have focused on the responses of dorsal raphe serotonin neurons in response to rewarding and punishing stimuli as well as the cues that predict them. The first of these studies published examined serotonin neurons while mice engaged in a Pavlovian reward task (Liu et al., 2014). In this task, mice received one of two odor conditioned stimuli that predicted, at some delay, either a reward or no reward. 65% of neurons demonstrated a change in activity between odor cue and reward delivery on rewarded trials, with both phasic and ramping patterns. Consistent with findings from monkeys and rats, subsets of these neurons were either significantly excited or inhibited during this period. Reward trial selectivity was more prominent in the identified serotonin neuron population than a randomly sampled dorsal raphe population.

From this work, it remained unclear if observed responses were related to salience, behavior, or value. The second study, from J.Y. Cohen and colleagues, added substantial clarity to this question using a few variants of a Pavlovian task with both rewarding and aversive outcomes. In the first task, odors predicting either reward or punishment were presented to the mice in

blocks of trials, with a short delay between the conditioned stimulus and the outcome. All 29 serotonin neurons recorded in this variant were task responsive. Similar to recordings of dorsal raphe in monkeys, almost all recorded neurons were phasically responsive to the aversive airpuff and about half were phasically excited by the reward-predicting stimulus. In a second variant of the task, a third block type was introduced, with an odor that predicted nothing. Phasic responses during the conditioned stimuli were a function of value for all 23 neurons recorded, with firing rates greatest for reward-predicting stimuli and smallest for punishment-predicting stimuli. These representations of value, in both variants of the task, were also present in the long-timescale tonic activity of half of neurons, differentiating each block type. With the tonic changes, interestingly, both positive and negative correlations were reported. These tasks cast doubt on the interpretation that serotonin activity is simply related to salience, but was still consistent with differences in behavior—mice reacted differently to each type of conditioned stimulus. To address this possibility, the authors designed two more variants of the task. In the first, three conditioned stimuli were presented in separate blocks, predicting small rewards, large rewards, or nothing. For all 13 neurons recorded, tonic activity varied monotonically with value. In the second additional variant, four block types contained odor stimuli predicting reward, nothing, airpuff, or aversive quinine. About a third of neurons displayed differences in tonic activity between blocks of the two aversive stimuli. These results of this work demonstrated a clear relationship between serotonin neuron activity, valued outcomes, and learned expectations about those outcomes.

The third study to record from optogenetically-identified dorsal raphe serotonin neurons studied the population in a task in which mice had to run back and forth across a track in order to harvest reward from each end in an alternating fashion (Li et al., 2016). Four main response profiles were observed. The first showed a ramping excitation of firing rates as the mouse approached the reward delivery port, then a strong burst of spikes at the time of the reward. The second type also ramped when the animal entered the reward zone, but whose firing rates diminished quickly once reward was consumed. The third type showed slow

ramping across approach, reward delivery, and for some seconds after. The last type was the converse of the third, showing inhibition of firing instead of excitation. Consistent with the previous study, serotonin neuron activity appeared to track expectations about rewards and responded to the rewards themselves. The same study also recorded population level calcium signals of serotonin neurons in various reward and punishment contexts. The population responded with increased transients to sucrose consumption, male-male interaction, male-female mounting, mouse-shaped object investigation, and surprise sucrose delivery. Decreases in this activity were seen in response to unexpected foot shocks.

A handful of other studies have also measured activity at the population level using genetically-encoded calcium fluorescence indicators. These studies showed similar results to single neuron recordings in addition to correlations with surprising outcomes, locomotion, and pupil diameter (Matias et al., 2017; Zhong et al., 2017; Ren et al., 2018; Seo et al., 2019; Cazettes et al., 2021). However, given the heterogeneity of the responses of individual neurons (including opposite changes in firing rate in response to the same stimulus) interpretations of population averages are limited. A potential strategy for navigating this diversity with this type of measurement is to limit the population by projection target. Retrograde viral transfection with genetic specificity revealed that orbitofrontal-cortex-projecting and central-amygdala-projecting serotonin neurons in the dorsal raphe responded similarly to sucrose consumption, but with opposite changes in response to a foot shock (Ren et al., 2018). Chemogenetic activation of the central-amygdala-projecting population exclusively enhanced fear conditioning and anxiety-like behavior.

The consistency across serotonin neuron recordings, even despite some variation in behavioral context, is striking. While demonstrably a heterogeneous population, responses to valued outcomes and the cues that predict them is apparent. These relationships, interestingly, are present on multiple timescales, potentially indicating a multiplexing of signals with distinct effects on downstream targets.

1.3 Serotonin neurons and learning

Information about valued outcomes as well as the cues and contexts that predict them could be useful in driving flexible behavior. Behavioral or cognitive (used interchangeably, hereafter "behavioral") flexibility is the ability of the brain to adapt appropriately to changes in the environment or internal state. Manipulation studies of serotonin implicate its activity in behavioral flexibility, in a manner consistent with a role in learning. Such a function is an intriguing possible explanation of serotonin function because it generalizes across various behaviors and is amenable to more precise definitions and quantification. Additionally, computations related to value would be useful in driving behavioral flexibility.

Some of the earliest studies to manipulate the serotonin system in behaving animals examined the role of the neuromodulator in responding to punishment. For example, pharmacologically elevating serotonin reduced rats' sensitivity to a mild shock when seeking reward (Wise et al., 1970) while depleting serotonin had the opposite effect (Wise et al., 1973). Results like these originally led to various hypotheses about the involvement of serotonin in anxiety and suppressing behavior in response to aversive outcomes (Soubrié, 1986). Extensive research has been conducted addressing these ideas and has been reviewed elsewhere (Soubrié, 1986; Cools et al., 2008). An alternative interpretation of these experiments is that animals' learning about the aversive outcomes was modulated by the manipulations. As a result, more recent research has examined serotonin neuron function in the context of fear learning (Bocchio et al., 2016; Sengupta and Holmes, 2019). Acute pharmacological elevation of serotonin in rats enhanced fear conditioning and expression, for example (Burghardt et al., 2004, 2007).

Learning from rewarding outcomes has also been an area of active research in the pursuit of understanding serotonin neuron function. Potentially similar to learning from aversive outcomes, the research suggests a role in learning when expected rewards are withheld. Observed effects from learning from rewards themselves are mixed. Work from Trevor

Robbins' lab provided some of the first explicit tests of related ideas. In one publication, rats were trained on a five-choice serial reaction time task before lesioning serotonin neurons by intracerebroventricular injection of 5,7-dihydroxytryptamine (Winstanley et al., 2004). In this task, rats had to pay attention to the lights in each of 9 separate nose-poke ports and were rewarded if they correctly poked into the port that had been briefly illuminated. If animals poked prematurely, the house lights would turn off and there would be a 5 second timeout period. Global depletion of serotonin (90% decrease in forebrain serotonin) did not affect correct responses to the lights, but did increase levels of premature responding. In addition to the proposed ideas about impulsive action and behavioral inhibition, these results are also consistent with a failure to learn from a no reward outcome.

Another study from the Robbins lab bidirectionally modulated serotonin in a probabilistic reversal learning task (Bari et al., 2010). Rats chose between two illuminated nose-poke ports that delivered reward probabilistically. The correct port delivered reward with probability 0.8 and the incorrect port with probability 0.2. The contingencies were reversed if the rat chose the correct port on 8 consecutive trials. In order to behave flexibly and maximize the amount of reward received, the rat must recognize the change in contingency and adapt their behavior accordingly. Serotonin was acutely elevated by administration of a selective serotonin reuptake inhibitor. The drug did not affect behavior in the acquisition phase and had dose dependent effects during reversal. In teasing apart effects on learning from rewards or no rewards, the researchers examined win-stay and lose-shift rates. These analyses calculate the probability of repeating the same choice after experiencing a reward (win-stay) and changing choices after receiving no reward (lose-shift). They found that low doses of the drug decreased the number of reversals completed as a consequence of increasing the lose-shift probability. Because rewards were probabilistic at the correct spout, changing choice frequently after a no reward outcome can lead to suboptimal behavior. Higher doses of the drug, however, increased the number of reversals completed by decreasing lose-shift rate. Lesioning serotonin neurons with 5,7-dihydroxytryptamine decreased the number of

completed reversals by both increasing lose-shift and decreasing win-stay probabilities. In both cases, these findings established a relationship between serotonin neuron activity and how rewarding outcomes, and lack thereof, are used to drive flexible reward behavior. The findings also show that serotonin neuron activity is not necessary for this learning, but may modulate the speed of it.

A similar learning deficit was observed in mice performing a Pavlovian reversal task using more precise methods of manipulation (Matias et al., 2017). In this task, mice were trained to associate specific cues with rewards, punishments, or nothing following a short delay. During the delay period, anticipatory licking of the mice indicated the values of the expected outcome. After a relatively long period of training, these associations were switched. Dorsal raphe serotonin neurons were selectively and reversibly inhibited using chemogenetics to express an inhibitory designer receptor exclusively in this population. The receptor is gated by an exogenous ligand that was administered when the associations were switched. This manipulation improves upon the temporally and spatially imprecise lesion methods previously used, which allow substantial time for compensatory mechanisms to take place and lesion serotonin neurons globally. Inhibition of serotonin neurons with chemogenetics impaired learning of the new contingencies, as evident in the slower adaptation of anticipatory licking in the time between the cue and outcome. Behavioral adaptation to outcomes that were less valuable than before was specifically impaired. Interestingly, bulk calcium fluorescence signals showed an increase in the activity of dorsal raphe serotonin neurons in response to the new, surprising outcomes. Taken together, these findings suggest that serotonin neuron activity may track violations in expectations in order to guide the rate of new learning.

The most explicit demonstration of the relationship between serotonin neuron activity and learning rate came in a reward-driven decision-making task in mice (Iigaya et al., 2018). In this restless two-armed bandit task, mice freely chose between two potential sources of rewards. The rewards were delivered probabilistically and the proababilities assigned to each source changed regularly. Similar to previously-described tasks, one source always had a

higher probability of reward delivery so the animals had to rely on their history of actions and outcomes in order to behave flexibly and maximize the rewards they received. Using quantitative models of behavior the authors showed that optogenetic activation of serotonin neurons in mouse dorsal raphe enhanced how quickly the animals learned from outcomes to drive this decision making behavior.

Results consistent with these studies have also been observed in humans. Trevor Robbins' lab has also conducted numerous experiments in humans in which the subjects undergo tryptophan depletion prior to behavior. Tryptophan is crucial to the synthesis of serotonin and its reduction leads to decreases in brain serotonin and release (Biggio et al., 1974). The first of many studies from the lab using this technique revealed an impairment in various visual discrimination tasks that required spatial memory (Park et al., 1994; Rogers et al., 1999). In one of these tasks, participants had to learn which feature of a compound visual stimulus made it the correct choice over a second one. The stimuli were shapes with lines overlaid, so the subject had to discern which of these two dimensions was informative and which feature of that dimension was correct (e.g., square or circle if shape was the relevant dimension). The rule was then reversed within the same dimension or across dimensions. In all types of reversals, tryptophan-depleted subjects were slower to adapt to the new rule.

In the most recent of these studies, effects of tryptophan depletion were examined in a two-alternative forced choice reversal task with positive and negative feedback (Kanen et al., 2020). When prompted with a certain visual stimulus, participants learned to make one of two responses. Correct responses resulted in either positive or neutral (lack of) feedback and incorrect responses resulted in neutral or negative feedback. The contingencies between stimulus and correct action were regularly reversed. The most consistent effect of tryptophan depletion was an impairment in adaptation following a reversal in correct-positive-, incorrect-neutral-feedback blocks. Despite the lack of positive feedback for a previously-correct choice, tryptophan-depleted subjects perseverated longer on the incorrect choice. Across individual subjects, the magnitude of the impairment correlated with the amount of tryptophan depletion.

The interpretation that serotonin is involved with learning from less-than-expected outcomes is consistent with these findings from humans.

The work reviewed here represents just a subset of serotonin activity manipulations in behaviors involving learning and cognitive flexibility (Roberts et al., 2020). Flexibility in social contexts, for example, has been linked to serotonin neuron function (Kiser et al., 2012; Dölen et al., 2013; Nardou et al., 2019). Results from studies of serotonin in other behavioral contexts could also be explained in the framework of learning. However, these alternative interpretations of data are beyond the scope of this introduction and will be expanded upon in discussion sections.

1.4 Theory of learning and decision making

In the manipulation experiments, animals are still capable of learning about the correlational relationships between stimuli, actions, and outcomes, but the speed at which they do so is affected by the manipulation. The hypothesis then arises that serotonin neurons *modulate* the learning process, possibly by regulating the rate at which learning occurs. Modulating learning rates is important for generating adaptive behavior in such a rich, noisy, and changing world. For example, if an environment—and the correlational relationships in it—change frequently, it is best to learn quickly, as old observations are more likely to be irrelevant. Conversely, if an environment is stable, but the correlational relationships are noisy, it is better to learn slowly. Slower learning in these circumstances prevents behavior from being driven sub-optimally by short-term fluctuations. During a morning commute, for example, if one experiences higher than normal levels of traffic one day, one will likely not change their route the next day because traffic is known to be variable.

Learning is a cognitive process that is not directly observable in behavior. In the commuting example, if one were to experience a week of elevated levels of traffic, they may change their route the following day. As an outside observer, how do we characterize the

learning that occurs each day of that week when there is no change in behavior? One approach is to describe in mathematical operations a proposed mechanism by which learning and subsequent decision making occurs. This combination of theory and computation has proved fruitful in constraining what types of latent cognitive processes could produce observable behavior.

In behavioral tasks that require sequential learning about correlational relationships, reinforcement learning algorithms provide an elegant model of learning and decision making. In one class of these models, the brain learns the value of an action by the outcomes that result from taking it (Figure 1-2). The brain can then compare the values of different actions in order to make a decision, with the overall goal of maximizing value. The action values are updated by comparing the expected value of that action with the actual outcome. The difference, known as the reward prediction error, updates the old expected value at some rate. This continuous process is very effective at maximizing reward and characterizing real behavior in certain contexts. There is even evidence that the brain implements something like reinforcement learning. The activity of dopamine neurons correlates with the reward prediction error (Schultz et al., 1997; Cohen et al., 2012) and neurons the medial prefrontal cortex track the relative values of available actions(Massi et al., 2018; Wang et al., 2018; Bari et al., 2019).

The math describing this learning and decision making process becomes clear through example. In each trial of a two-armed bandit task, an agent has the choice between two actions, left and right, and wants to maximize reward received by choosing the option that is more rewarding. The amount of reward that results from each action is initially unknown to the agent. Assuming the agent chooses the left option, the reward prediction error, δ, is calculated as follows:

$$\delta(t) = R(t) - Q_l(t),$$

where R is the amount of reward received, Q_l is the expected value of making a left action,

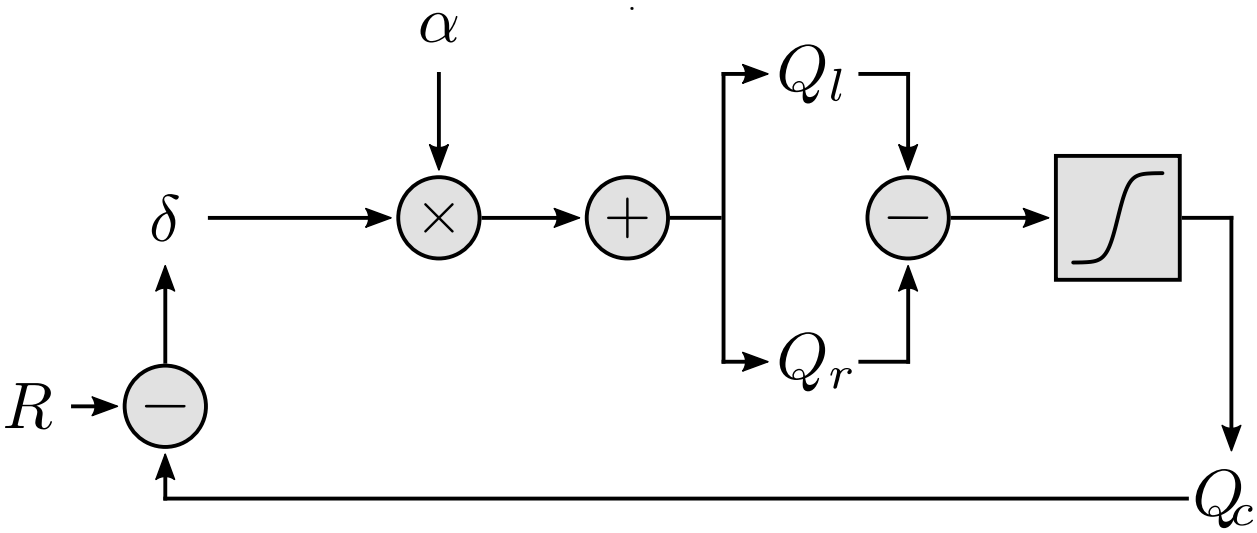

Figure 1-2: **Action value reinforcement learning.** Schematic of the model algorithm. *Relative value* $(Q_r - Q_l)$ is used to make choices by the decision function denoted by the sigmoidal, softmax function. The predicted value of a choice (Q_c) is compared to reward (R), to generate a *reward prediction error* (δ). δ is used to update Q_c at a rate controlled by α.

and t denotes the trial at which that action takes place. The agent's expected value of that action is updated according to the following equation:

$$Q_l(t+1) = Q_l(t) + \alpha\delta(t)$$

in which the parameter α is the rate at which learning occurs. If the reward is less than expected, the negative reward prediction error will diminish the new expected value and *vice versa*. In this way, the expected value will eventually converge to the actual value of the reward. The action values are then fed into a decision function in order to select an action on the next trial. One such function is the softmax (Daw et al., 2006):

$$P(c(t) = r) = \frac{1}{1 + e^{-\beta(Q_r(t) - Q_l(t))}},$$
$$P(c(t) = l) = 1 - P(c(t) = r),$$

where the probability of choosing the right spout at trial t is given by $P(c(t) = r)$. The equation describes a sigmoidal function in which the probability of choosing an action increases as a function of how much greater that action's value is than the other. The inverse temperature parameter, β controls the steepness of this function: a higher β means that

the same difference in action values results in a larger choice probability for the action with greater expected value. In other words, β controls how stochastic the behavior is relative to the expected values.

Reinforcement learning models such as the one described, as with all models, fail to describe certain features of observed behavior. One limitation of these models is that the learning rate parameter, α, is often treated as a fixed value. Learning rates, however, are demonstrably variable. Behrens et al. showed that increasing environmental volatility enhanced learning rate in humans (Behrens et al., 2007). In this one-armed bandit task, the contingencies between actions and outcomes were probabilistic with one option always having a higher reward probability than the other. The reward probability assignments were changed at different rates throughout the task, resulting in stable periods and volatile ones. The study found that volatile periods engendered faster learning rates. From a normative perspective, this is an optimal strategy since only recent outcomes are relevant to the value of an action when the contingency between action and outcome changes frequently.

Volatility is an objective measurement of the environment, so what information or estimates could the brain be using to modulate learning rates? What are the computations involved? In another behavioral assay, M.R. Nassar and colleagues sought to model how learning rates were being modulated to behave adaptively in a changing environment (Nassar et al., 2012). In their predictive inference task, human subjects tried to predict a number that was drawn from a Guassian distribution with varying noise, and whose mean changed at different frequencies throughout the session. The researchers found that behavior could be well described by an optimal Bayesian model that used uncertainty and change point predictions to modulate learning. Specifically, when a change point in the mean of the outcomes was detected, learning rates would increase because old outcomes were no longer relevant to the current, presumably new, mean. When the estimate of the mean is uncertain, detecting change becomes more difficult. Uncertainty arises from noise in the generative distribution as well as insufficient experience with that distribution. The latter type will increase learning rate, since it is

important to learn quickly about new circumstances. The former, however, will decrease learning rate because behavior should not be driven by short-term fluctuations in outcomes if the generative distribution is noisy, but stable. One should not change their belief that the probability of a fair coin landing heads side up is 0.5 just because they observed it landing tails side up 10 times in a row, for example. The model captured the human tendency to increase learning rates at change points. Further, pupil diameter changes correlated with the estimated change-point probability and average pupil diameter tracked uncertainty. Interestingly, recent work showed that optogenetic activation of dorsal raphe serotonin neurons in mice increased pupil size and that the magnitude of this effect was modulated by uncertainty (Cazettes et al., 2021).

Normative models, like the Bayesian inference model described above, prescribe how the brain should optimally adapt to environmental conditions and their dynamics under uncertainty (Yu and Dayan, 2005; Payzan-LeNestour and Bossaerts, 2011; Nassar et al., 2012; O'Reilly, 2013; Payzan-LeNestour et al., 2013; McGuire et al., 2014; Faraji et al., 2018). These models, however, assume knowledge of the structure of the environment and are computationally intractable in neural structures. The brain is capable of learning structure in the environment but it must be doing so in a more computationally efficient manner. Thus, it is important to find appropriate approximations to these models in order to construct hypotheses about their neural implementation. Some theories that have tried to address this issue have roots in economic ideas about risk and have shown applicability to behavior and neural function (Preuschoff et al., 2006; Preuschoff and Bossaerts, 2007; Preuschoff et al., 2008). In one of these theories, learning rates should be equivalent to the projection coefficient of expectations onto past prediction errors (Aoki, 1987; Preuschoff and Bossaerts, 2007). This coefficient was shown to depend on the covariance between expectations and reward prediction errors that are scaled by the variance of prediction errors (Preuschoff and Bossaerts, 2007). The covariance determines learning such that accurate expectations in low variability environments will enhance it. However, calculation of covariance requires the

maintenance of individual previous expectations and scaled errors—they cannot be updated iterative fashion—which places a significant load on working memory. This model also did not specify how the variance of prediction errors was computed.

The brain faces a difficult challenge in learning correlational relationships between stimuli or actions and outcomes, especially when those relationships are noisy and subject to change. In noisy, dynamic environments these models leverage uncertainty to modulate how information is incorporated into beliefs; in other words, uncertainty should be used to drive learning rates.

1.5 Theory of serotonin neuron function

A few groups have proposed mechanistic theories of serotonin function. Kenji Doya presented one of the first ideas, speculating that serotonin was responsible for temporal discounting (Doya, 2002). Thus far the work described in this introduction treats the outcomes of actions as discrete events. In reality, the consequences of an action are often temporally extended. When a high school student completes a homework assignment, they do not only gain the value of getting a better grade in the course, but they increase the value of their state; there is the added value of increasing their chance to be admitted to their ideal university. Known as temporal discounting, humans discount the value of outcomes that occur farther in the future. High temporal discounting can ease the computational burden of decision making. Otherwise, that student would be stuck contemplating if a sip of water or a sip of coffee right now will better their chances of university admission. Conversely, the choice of which extracurricular activity to pursue may warrant less temporal discounting. Consistent with this theory, work from Kenji Doya shows that optogenetic stimulation of dorsal raphe serotonin neurons increases the amount of time that mice will wait for a temporally-uncertain reward (Miyazaki et al., 2014). However, the experiments cannot rule out other explanations. The activation of serotonin neurons may be disrupting the animals' perception of time, impairing their movement (Correia et al., 2017), causing a distracting sensation, or impairing how they

learn from no rewards as time passes. Further, the activity of serotonin neurons has not been measured in this behavioral context.

Others have proposed a role for serotonin in the learning process. Nathaniel Daw, Sham Kakade, and Peter Dayan theorized that serotonin modulated learning through its opposition to dopamine (Daw et al., 2002). In their modification of a standard reinforcement learning model, serotonin estimated the overall reward rate per the following equation:

$$\bar{R}(t+1) = \bar{R}(t) + \alpha_{\bar{R}}(\bar{R}(t) - R(t)),$$

where $\bar{R}$ is the average reward rate and $\alpha_{\bar{R}}$ controls how quickly this moving average is updated. Returning to the two-armed bandit example, if an agent makes a left choice the reward prediction error is calculated by

$$\delta(t) = R(t) - Q_l(t) - \bar{R}(t).$$

Here, the average reward is subtracted from the typical prediction error. In this formulation, immediate outcomes are weighed against long-term average reward. High average reward, for example, amplifies the effects of negative reward prediction errors on learning and mitigates positive errors. Much like Doya's theory, the opponency model was put forth to generate testable hypotheses, but was based mostly on indirect pharmacological manipulations. In particular, they reference studies showing that activating the serotonin system had opposite effects on conditioned behaviors as dopamine activation (Fletcher et al., 1993, 1995; Fletcher and Korth, 1999; Fletcher et al., 1999).

One recent paper ventured to test the opponency hypothesis explicitly (Wittmann et al., 2020). Monkeys were trained on a type of bandit task in which they chose between two of three possible options on a given trial. The outcomes, a reward or nothing, were determined probabilistically for each choice. Halfway through the session these contingencies switched. To characterize behavior they modified the opponency model slightly to add a parameter that determines how average reward affects reward prediction errors (global reward state model):

$$\delta(t) = R(t) - Q_l(t) + \omega\bar{R}(t),$$

where ω controls both the sign and the magnitude of the effect. When fit to behavior, ω was a positive value, in contrast to the opponency model. The behavioral evidence for this model, in addition to BIC score, was related to the relationship between reward history and win-stay, lose-shift analyses. In the actual data and simulation, they show that when reward history is high, win-stay rate is higher and lose-shift rate is lower. However, this phenomenon can be recapitulated with a standard reinforcement learning model. Using fMRI observations of BOLD signals, the study also showed a negative correlation between the average reward term in their model and activity in the dorsal raphe. This mode of measurement is limited by spatial and temporal resolution, recording the population average, slow timescale activity of all dorsal raphe neurons. There were also no causal manipulations to test the predictions of the model when the serotonin system was altered.

Given these limited results, the validity of these models remains to be tested. However, behavioral evidence and reasoning from first principles seems to cast doubt on their applicability. In the aforementioned study from Nassar et al., learning rates increased when a change in the environment was detected (Nassar et al., 2012). Detecting change is easier when that change is more obvious. In the language of their model, easy change detection would occur when the subject is confident of the current value through repeated experience, there is low noise in that value, and the change in value to the new one is large. In this case, learning rate would be very high when an action that was previously always rewarded suddenly stops producing reward. Since reward rate would be high prior to the change, the global reward state predicts the opposite, in contrast to observed behavior. When the contingencies between action and outcome are noisy (e.g., the probability of the best action being rewarded is 0.5) but stable, the ideal model shows that learning rates should be low. However, in this case of the opponency model predicts median learning rates. In terms of producing optimal behavior, normative models suggest that uncertainty should drive learning.

1.6 Prefrontal cortex, learning, and uncertainty

In addition to behavioral evidence, there are observations of neural activity that indicate that the brain uses uncertainty in modulating how much to learn. Some of these findings implicate the prefrontal cortex in the process. The prefrontal cortex has been the subject of intense focus over the years, particularly in the context of flexible behavior. Lesions (Kennerley et al., 2006) and pharmacological inactivation (Shima and Tanji, 1998; Bari et al., 2019) of prefrontal cortex impair goal-directed behavior that is dependent on the continuous maintenance of the representations of action values. Indeed, activity of single neurons in prefrontal cortex track the value of choices even as those values change (Bari et al., 2019). As monkeys made choices in a two-armed bandit task with changing outcome values, prefrontal cortex representations of object value were updated by experienced rewards that followed choices of that object (Tsutsui et al., 2016). A similar observation was made in mice behaving in a two-armed bandit task with probabilistic and dynamic reward contingencies (Bari et al., 2019). Single neuron firing rates persistently encoded the relative values of the two available choices over long timescales. These relative values were updated according to experienced outcomes in a manner consistent with a proposed reinforcement learning model. The maintenance of such values over the long periods of time between trials are thought to be useful for subsequent decision making. Indeed, prefrontal cortex activity predicted probability of choice and was shown to be necessary for flexible behavior using pharmacological inhibition.

Previously described studies of behavior demonstrated that the rates at which learning occurs vary in response to the statistics of the environment (Behrens et al., 2007; Nassar et al., 2012). It follows then, that how quickly representations of value are updated should follow the same patterns. Manipulations of environmental volatility in a monkey two-armed bandit task seem to support this idea (Massi et al., 2018). When contingencies between action and outcome changed frequently (high volatility), outcome signals in the orbitofrontal cortex were enhanced. Joint choice and outcome encoding in the dorsolateral prefrontal

cortex was also enhanced in high volatility blocks. These changes in neural activity coincided with faster learning rates in observed behavior.

In addition to objective statistics of the environment, brain activity related to learning was also modified by estimates of uncertainty and change point probability (McGuire et al., 2014). These findings relied on the ideal Bayesian model described above to generate these theoretical latent cognitive variables (Nassar et al., 2012). In the dorsomedial prefrontal cortex, posterior cingulate cortex, and right lateral prefrontal cortex, they found correlations between fMRI BOLD signals and change point probability, uncertainty, and responses to reward. This convergence of information suggests that these regions are involved in adaptive learning. A more recent study from the same group replicated these findings and expanded upon them using a different experimental design (Kao et al., 2020). The authors found that posterior cingulate cortex activity correlated with error magnitude that was modulated by environmental uncertainty. Further, activity in the orbitofrontal, anterior cingulate, dorsomedial prefrontal, and dorsolateral prefrontal cortices predicted choice behavior as a function of error magnitude. These findings bolster the hypothesis that prefrontal cortex is involved in value computations that involve variable learning as a function of uncertainty.

1.7 Serotonin in prefrontal cortex

Given that the prefrontal cortex is such a crucial node for flexible decision making and tracks values as they are updated, the region is a prime candidate to mediate potential effects of serotonin on learning rates. The prefrontal cortex is the target of dense axonal arborizations from dorsal raphe serotonin neurons (Jacobs and Azmitia, 1992; Linley et al., 2013; Prouty et al., 2017; Ren et al., 2018). Serotonin receptor expression in these regions is dense and, along with other regions associated with learning and decision making, accounts for the majority of the 5-HT_2 family of receptors in the central nervous system.

Serotonin has interesting effects on prefrontal cortex neurons and circuits that suggest

the ability of the neurotransmitter to modulate what and how information is processed there. Support for this idea comes from differential effects of serotonin on different neuron subtypes in the mouse medial prefrontal cortex observed during whole cell recordings *in vitro*. Neurons defined by their callosal/commissural projections showed excitatory or biphasic (excitation followed by inhibition) responses in response to bath application of serotonin (Avesar and Gulledge, 2012; Stephens et al., 2014). The excitatory component was mediated by 5-HT_{2A} receptors. Corticoamygdalar neurons showed similar response profiles mediated by the same receptor, but the excitatory effects were contingent on extrinsic excitatory drive (Avesar et al., 2018). Corticopontine neurons, on the other hand, were all inhibited by serotonin via 5-HT_{1A} receptors. These projection-pathway-specific modulations of prefrontal cortex by serotonin provide a mechanism by which outgoing information can be regulated.

In addition to outputs, inputs to medial prefrontal cortex are differentially gated by serotonin. In one study, recordings from layer 5 pyramidal neurons were made while optogenetically stimulating inputs from various regions (Kjaerby et al., 2016). Bath application of serotonin decreased the magnitude of monosynaptic excitatory inputs from contralateral medial prefrontal cortex and ventral hippocampus while sparing those from mediodorsal thalamus. The inhibitory effects of serotonin were mediated by presynaptic 5-HT_{1B} receptors. Infusion of a 5-HT_{1B} agonist in medial prefrontal cortex *in vivo* resulted in a suppression of theta frequency oscillatory activity—a population average of activity, the frequency of which was previously associated with hippocampal inputs (Adhikari et al., 2010, 2011). Together these results demonstrate the serotonergic modification of local circuit activity at least in part due to modulation of inputs to the region. Circuit-level effects in medial prefrontal cortex are also mediated by effects of serotonin on fast-spiking interneurons (Athilingam et al., 2017). Serotonin depolarized these neurons through 5-HT_{2A} receptors and affected passive membrane properties. Firing in response to gamma frequency inputs was preferentially enhanced through the serotonin-mediated slowing of excitatory potential decay. These findings provide another mechanism by which serotonin can mediate information processing in prefrontal cortex by

biasing circuits towards certain patterns of activity.

As the anatomy and electrophysiology would imply, serotonin has a functional role in prefrontal cortex during behavior. Infusion of a $5HT_{1A}$ agonist or $5HT_{2A}$ antagonist into medial prefrontal cortex improved accuracy and decreased premature responding in a five-choice serial reaction time task, respectively (Winstanley et al., 2004). Infusion of a $5\text{-}HT_{1B}$ agonist into the same region reduced avoidance of an anxiogenic region of an elevated plus maze (Kjaerby et al., 2016). Optogenetic stimulation of serotonin neuron axons in orbitofrontal cortex following a stimulus was sufficient to entrain an expectation-like response to the stimulus (Zhou et al., 2015).

Some of the clearest behavioral effects of serotonin manipulation were revealed using lesions of serotonin axons in the prefrontal cortex of rhesus monkeys (Clarke et al., 2004, 2007). In this deterministic reversal learning task, monkeys learned to discriminate pairs of visual stimuli, one that resulted in reward and one that did not. In the first study, the monkeys learned two pairs of discriminations before serotonin axons in prefrontal cortex were lesioned with 5,7-dihydroxytryptamine. Monkeys were tested on retention of one of the original pairs, acquisition of a new pair, and 4 reversals of that new pair. The lesions impaired the animals' ability to adapt their behavior after a reversal, as they continued to choose the previously-rewarded stimulus. In the second study, the authors sought to disambiguate effects of learned avoidance of the previously-nonrewarded stimulus and perseveration on the previously-rewarded stimulus. After the first reversal, they split the two stimuli into pairs with new stimuli. Serotonin-lesioned animals were able to successfully choose the previously-nonrewarded stimulus in its new pairing. Interestingly, animals perseverated on the previously-rewarded stimulus in the other despite no longer receiving a reward for their choice. This selective impairment using a keen modification of the behavior assay suggests that serotonin in prefrontal cortex is important for learning from less-than-expected outcomes. It is important to note again that animals were still capable of learning, but did so more slowly. This result is consistent with a role for serotonin neurons in modulating the rate of

learning in order to drive flexible behavior.

These findings were recapitulated in rats (Alsiö et al., 2021). Lesions of serotonin axons in orbitofrontal cortex using 5,7-dihydroxytryptamine impaired animals ability to adapt to a change in choice-outcome contingencies. Lesions in medial prefrontal cortex left reversal behavior intact, but impaired the initial acquisition of choice-outcome contingencies.

1.8 Research motivation

The serotonin literature is vast and this introduction has just skimmed the surface, only covering the findings relevant to the research that follows in this dissertation. Despite the extensive body of work, many questions remain about serotonin neuron function. This is, in part, due to the dearth of electrophysiological recordings of identified serotonin neurons in awake, behaving animals. The few recordings of this kind implicate a role for serotonin neurons in processing valued outcomes and the cues and contexts that predict them. Manipulations of serotonin neuron activity result in changes in behavior consistent with changes in learning rate. While some have proposed theories of serotonin neuron computation, they have not been explicitly tested. Thus, the computations of serotonin neurons, that link neural activity, cognition, and behavior are yet unknown.

Understanding these computations related to value, learning, and decision making will benefit from observations of adaptive behavior in dynamic environments and computational models of that behavior. In these contexts, varying the statistics of relationships between cues or actions and outcomes can distinguish potential cognitive processes and their implementation by serotonin neurons. Precise manipulations of the activity of these neurons can test the predicted changes in behavior made by these models. Model predictions can be tested further by investigating the effects of the serotonin signal in downstream regions involved in value computations. While the research presented in this dissertation is very far from a definitive and comprehensive explanation of serotonin neuron function, it follows the above strategy

in order to propose one possible mechanistic explanation of that function in the context of adaptive, reward-motivated behavior.

1.9 Disclosures

Two manuscripts were reformatted in adherence to the requirements of The Johns Hopkins University for this dissertation. The first is contained within *Chapter 2* and *Chapter 3* (Grossman et al., in submission). The second is contained in *Chapter 4* (Grossman et al., in preparation).

Chapter 2

Uncertainty modulates learning rate in a mouse model of dynamic foraging

Abstract

The rate at which learning occurs should not be static. Rather, it should be modulated to complement the statistics of an environment. When the relationship between an action and outcome is probabilistic, but stable, learning should be minimized. This mitigation prevents behavior from being influenced by short-term fluctuations in outcomes due to the stochastic nature of the relationship. However, when a change in that relationship occurs, learning should be enhanced in order to drive flexible behavior. One approach to dealing with the difficulty of detecting change in noisy environments is to leverage an estimate of uncertainty about the action-outcome contingency. This expected uncertainty tracks variability in the contingency and down-regulates learning. Unexpected uncertainty then tracks deviations from this expected variability, and up-regulates learning. Here, we show that mice behaving in a dynamic foraging task can be understood to be using such a strategy to harvest rewards. A model that incorporates uncertainty to drive learning in this way was able to explain decision making behavior in the foraging task as well as a Pavlovian version of the task. The model was also able to capture features of foraging behavior that a model with static learning rates could not.

2.1 Introduction

Models from control theory and reinforcement learning (RL) propose that behavioral policies are learned through interactions between the nervous system and the environment (Bertsekas and Tsitsiklis, 1996; Sutton and Barto, 1998). Within this framework, an animal learns from discrepancies between expected and received outcomes of actions (reward prediction errors, RPEs). The rate at which learning occurs is usually treated as a fixed parameter, but optimal learning rates vary when the environment changes. Consequently, animals should vary how rapidly they learn in order to behave adaptively and maximize reward. Normatively, learning rates should vary as a function of uncertainty (Dayan et al., 2000; Yu and Dayan, 2005; Soltani and Izquierdo, 2019). When some amount of uncertainty is expected (also referred to as outcome variance or risk), learning rates should decrease (Preuschoff et al., 2006; Preuschoff and Bossaerts, 2007; Preuschoff et al., 2008; Diederen and Schultz, 2015). Slower learning helps maximize reward when relationships between actions and outcomes are probabilistic but stable. This prevents animals from abandoning an optimal choice due to short-term fluctuations in outcomes. However, it is also important to detect changes in the underlying statistics of an environment. Here, deviations from expected uncertainty ("unexpected uncertainty") should increase learning rates (Yu and Dayan, 2005; Payzan-LeNestour and Bossaerts, 2011; Payzan-LeNestour et al., 2013; O'Reilly, 2013; Faraji et al., 2018). Tuning decision making in this way is known as "meta-learning", and there is evidence that humans and other animals use this strategy (Behrens et al., 2007; Herzfeld et al., 2014; Massi et al., 2018; Soltani and Izquierdo, 2019). It is yet unclear if mice engage in meta-learning. The exact computations by which the brain uses the statistics of the environment to guide learning are also unknown. We developed a dynamic foraging task with variable reward statistics for mice in order to address these questions.

2.2 Results

Mice behave adaptively in a dynamic foraging task

We trained thirsty, head-restrained mice (21 female, 27 male) on a dynamic foraging task in which they made choices between two alternative sources of water (Bari et al., 2019). Sessions consisted of about 300 trials (280 ± 66.6) with forced inter-trial intervals (1–31 s, exponentially distributed). Each trial began with an odor "go" cue that informed the animal that it could make a choice, but otherwise gave no information (Figure 2-1a,b). During a response window (1.5 s) the mouse could make a decision by licking either the left or the right spout. As a consequence of their choice, water was delivered probabilistically from the chosen spout. The reward probabilities ($P(R)$) assigned to each spout changed independently and randomly, in blocks of 20–35 trials (drawn from a uniform distribution). These reward contingencies were drawn from a set of three probabilities (either $P(R) \in \{0.1, 0.5, 0.9\}$ or $P(R) \in \{0.1, 0.4, 0.7\}$ for a given mouse) and were not signaled to the animal.

Mice mostly chose the higher-probability spout (Figure 2-2a; correct rate, 0.68 ± 0.029) and harvested rewards (reward rate, 0.57 ± 0.021 rewards trial^{-1}) over many sessions (13.8 ± 6.93 sessions mouse^{-1}). We first fit statistical models to quantify the effect of outcome history on choice. These logistic regressions revealed that mice used experience of recent outcomes to drive behavior (Figure 2-1c; time constants, 1.31 ± 0.25 trials for rewards, 1.04 ± 0.13 trials for no rewards, 95% C.I.). Similarly, we quantified the effect of outcomes on the latency to make a choice following the go cue. Consistent with previous findings (Bari et al., 2019), this model demonstrated a large effect of recent rewards on speeding up response times (Figure 2-2b,c; time constant, 1.76 ± 0.27 trials, 95% C.I.).

Mouse learning is not static

These statistical findings indicate that mice dynamically learned from recent experience. To understand the nature of this learning, we constructed a generative model from a family of

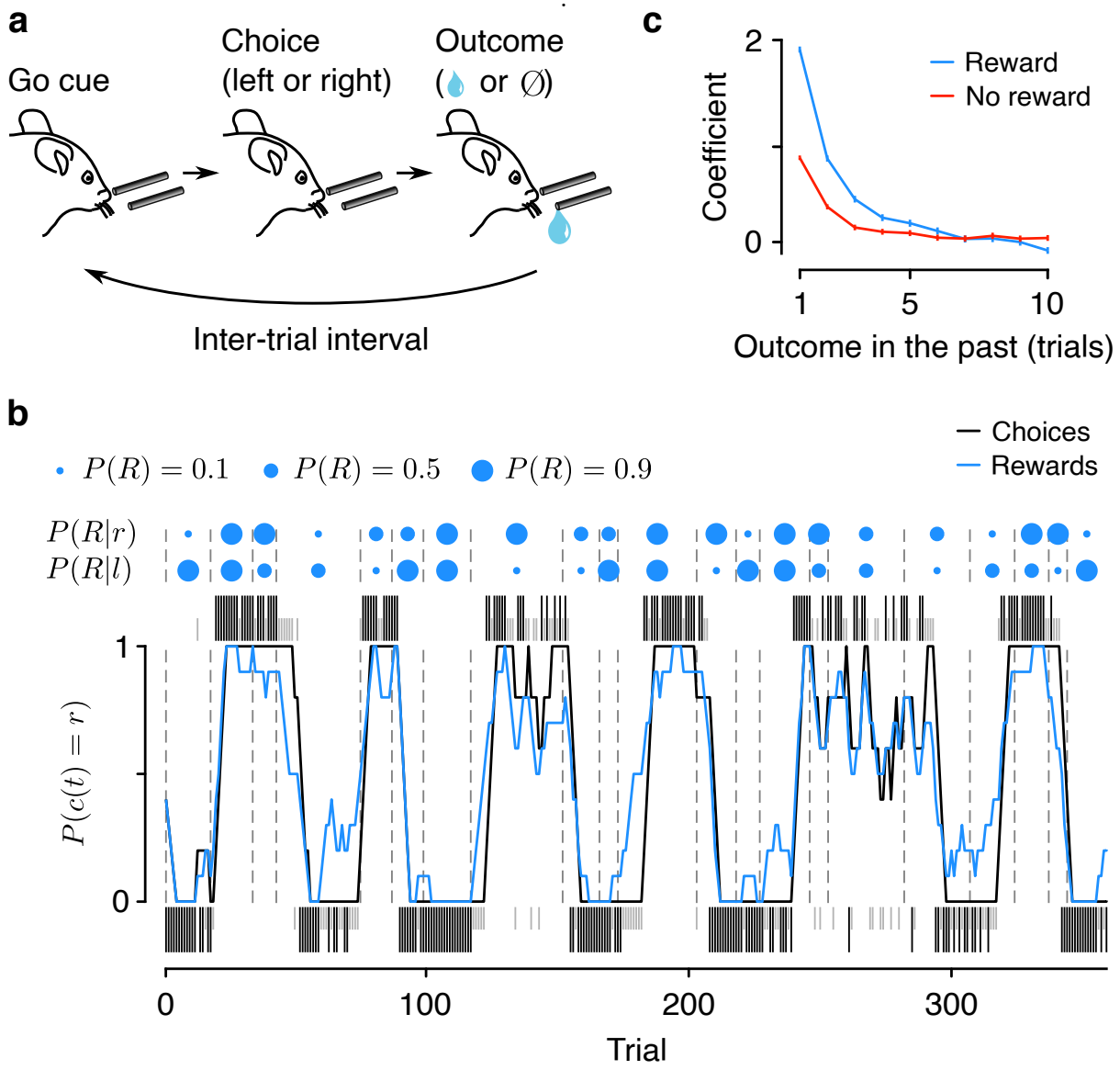

Figure 2-1: **Mice forage dynamically for rewards.**(a) Dynamic foraging task in which mice chose freely between a leftward and rightward lick, followed by a reward with a probability that varied over time. (b) Example mouse behavior from a single session in the task. Black (rewarded) and gray (unrewarded) ticks correspond to left (below) and right (above) choices. Black curve: mouse (smoothed over 5 trials, boxcar filter) choices. Blue curve: rewards (smoothed over 5 trials, boxcar filter). Blue dots indicate left/right reward probabilities and dashed lines indicate a change in reward probability ($P(R)$) for at least one spout. (c) Logistic regression coefficients for choice as a function of outcome history. Error bars: 95% CI.

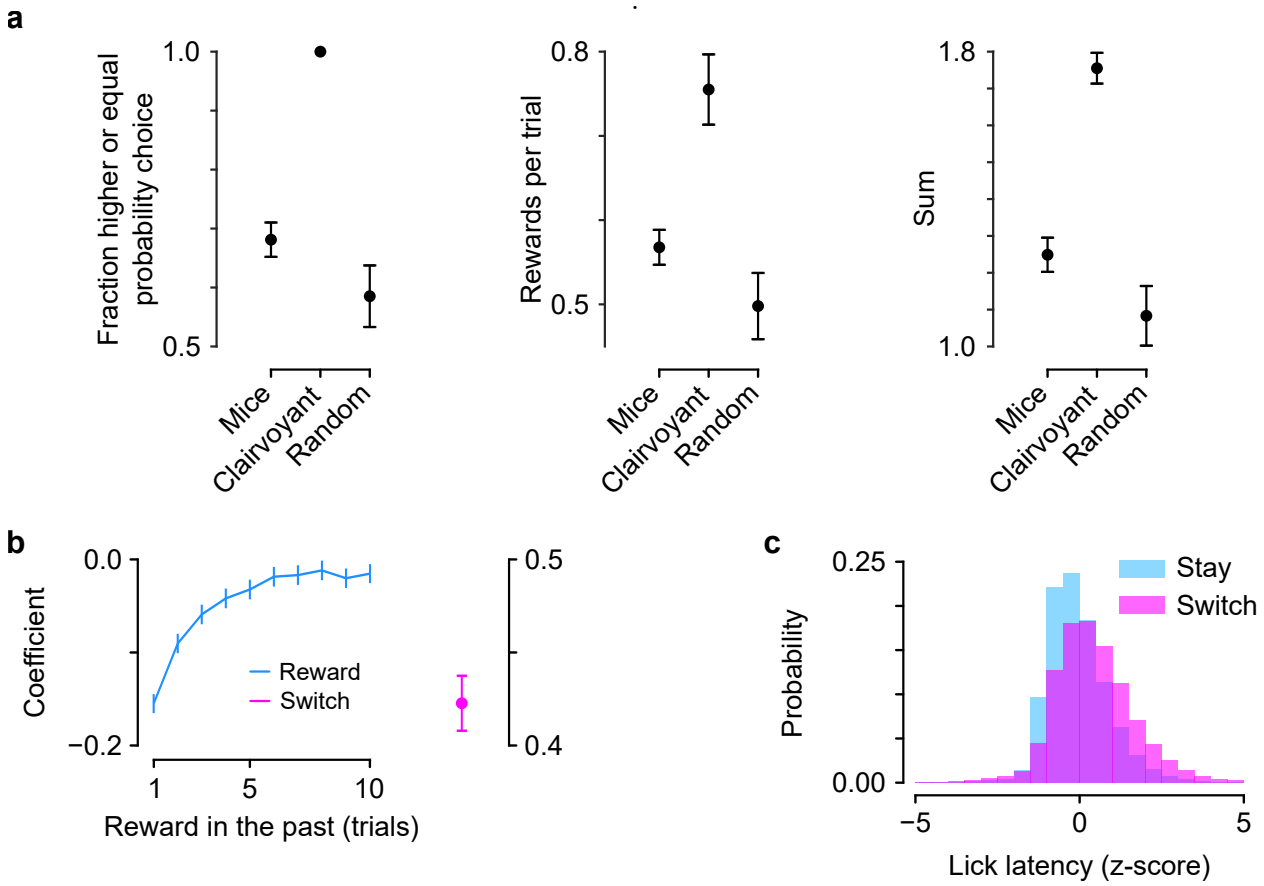

Figure 2-2: **Mice successfully harvest rewards and response time reflects reward history.** (a) Fraction of higher-probability choices, rewards per trial, and the sum of these quantities for mice, a "clairvoyant" model that knew reward probabilities, and random choices (paired t-test between mice and random: higher-probability choice, $t = 11.11$, $p < 10^{-18}$; rewards per trial, $t = 10.87$, $p < 10^{-17}$; sum, $t = 12.30$, $p < 10^{-20}$). (b) Linear regression coefficients of response time on reward history. Coefficient for switch trials was included in the regression. (c) Lick latency was faster on trials in which mice repeated the same choice ("stay") compared to when they made a different choice ("switch"; paired t-test, $t_{47} = -12.85$, $p < 10^{-16}$).

RL models called Q-learning (Bertsekas and Tsitsiklis, 1996; Sutton and Barto, 1998). This class of models creates a behavioral policy by maintaining an estimate of the value of each action (the expected reward from making that action). Using these values to make choices, the model then learns from those choices by using the RPE to update the action values, thereby forming a new policy (Figure 2-3a). How much to learn from RPEs is determined by the learning rate parameters. While these parameters are typically fixed across behavior, they need not be constant; they could vary according to statistics of the environment (Behrens et al., 2007). This type of adaptation is known as meta-learning.

We first fit a model to mouse behavior in which learning rates were constant. The model included separate parameters for learning from positive and negative RPEs because learning from rewards and no rewards was demonstrably asymmetric (Figure 2-1c), consistent with previous reports (Lefebvre et al., 2017; Dorfman et al., 2019; Dabney et al., 2020). This model fit overall behavior well (Bari et al., 2019), but was unable to capture a specific feature of behavior around transitions in reward probabilities (Figure 2-3b). In rare instances, both reward probabilities were reassigned within 5 trials of each other. When the probability assignments flipped from high and low to low and high (for example, from 0.9 on the left and 0.1 on the right to 0.1 on the left and 0.9 on the right), mice rapidly shifted their choices to the new higher-probability alternative. However, when reward probabilities transitioned from medium and low to low and high (for example, from 0.5 on the left and 0.1 on the right to 0.1 on the left and 0.9 on the right), mice took longer to adapt to the change (Figure 2-3b; effect of trial from transition $F_{1,28} = 217$, $p < 10^{-13}$ and trial from transition $\times$ transition type interaction $F_{1,28} = 5.23$, $p = 0.030$, linear mixed effects model). This difference in choice adaptation was still apparent when choice probabilities prior to the transition were identical (Figure 2-4a), demonstrating that the difference in outcome history is responsible for this effect on choice adaptation.

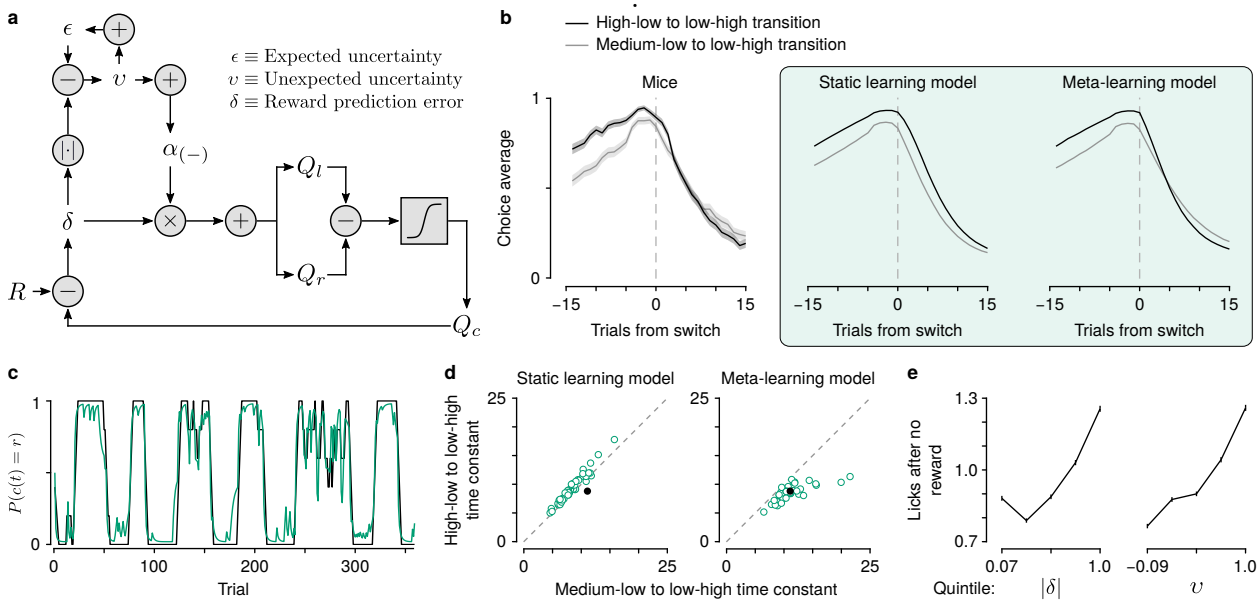

Figure 2-3: **Mice learn at variable rates.** (a) Schematic of the meta-learning model algorithm. *Relative value* ($Q_r - Q_l$) is used to make choices. The predicted value of a choice (Q_c) is compared to reward (R), to generate a *reward prediction error* (δ). $|\delta|$ is used to update *expected uncertainty* (ϵ), which is compared with the prediction error to generate *unexpected uncertainty* (υ). υ then determines how rapidly to learn from δ, thereby updating Q_r and Q_l. (b) Left: actual mouse behavior at transitions in which reward probabilities change simultaneously ($n = 225$ high-low to low-high, $n = 236$ medium-low to low-high). Lines are mean choice probability relative to the spout that initially has the higher probability, shading is Bernoulli SEM. Middle: simulated behavior at transitions using static learning model parameters fit to actual behavior. Right: simulated behavior at transitions using meta-learning model parameters fit to actual behavior. (c) Estimated choice probability of actual behavior (black, same as Figure 1b) and choice probability estimated with the meta-learning model (green) smoothed over 5 trials (boxcar filter). (d) Time constants from exponential curves fit to simulated choice probabilities (like those shown in (b)) for each mouse ($n = 48$, green circles) compared to the actual mouse behavior (black circle). Left: static-learning model (probability that mouse data come from simulated data distribution, $p < 10^{-6}$). Right: meta-learning model ($p = 0.89$). (e) Spout licks following no reward as a function of $|\delta|$ from the static learning model (left, regression coefficient = 0.38, $p < 10^{-20}$) or ϵ from the meta-learning model (right, regression coefficient = 0.45, $p < 10^{-20}$).

Mouse learning can be characterized by meta-learning

Based on outcome history, the transition from high to low is more obvious than the transition from medium to low. This observation is consistent with learning rates varying as a function of how much outcomes deviate from a learned amount of variability (expected uncertainty). Thus, we designed a model (Figure 2-3a) that learns an estimate of the expected uncertainty of the behavioral policy by calculating a moving, weighted average of unsigned RPEs (Soltani and Izquierdo, 2019). Increases in expected uncertainty cause slower learning. This computation helps maximize reward when outcomes are probabilistic but stable (Dayan et al., 2000; Preuschoff and Bossaerts, 2007; Diederen and Schultz, 2015). The model then calculates the difference between expected uncertainty and unsigned RPEs (unexpected uncertainty), integrating over trials, to determine how quickly the brain learns from those outcomes (Krugel et al., 2009; Payzan-LeNestour and Bossaerts, 2011; Payzan-LeNestour et al., 2013; Faraji et al., 2018). Intuitively, large RPEs that differ from recent history carry more information because they may signal a change in the environment and should therefore enhance learning.

When we modeled the mouse behavior with meta-learning in this way, simulations using fitted parameters reproduced the transition behavior (Figure 2-3b-d). It was only necessary to modulate learning from negative RPEs to capture the behavior of mice around these transitions, perhaps due to the asymmetric effect of rewards and no rewards on behavior (Figure 2-3b). Interestingly, not all forms of meta-learning were capable of mimicking mouse behavior. We were unable to reproduce the observed behavior using a model previously proposed to modulate learning rates and explain serotonin neuron function (Figure 2-4c,d; Daw et al., 2002; Wittmann et al., 2020). A Pearce-Hall model (Pearce and Hall, 1980), which modulates learning as a function of RPE magnitude in a different way, was also unsuccessful (Figure A2-4c,d).

To capture this transition behavior, our meta-learning model leveraged a higher learning rate following high-low to low-high transitions than following medium-low to low-high. Prior

to the transitions, expected uncertainty was lower when the animal was sampling the high probability spout as opposed to the medium probability spout (Figure 2-4e, $t_{473} = 11.8$, $p < 10^{-27}$, paired t-test). When the reward probabilities changed, the deviation from expected uncertainty was greater when high changed to low ($t_{473} = -7.78$, $p < 10^{-13}$, paired t-test), resulting in the faster learning rate ($t_{473} = -7.91$, $p < 10^{-13}$, paired t-test). We also looked at the dynamics of the latent variables within blocks to see if they evolved on timescales relevant to behavior and task structure. While block lengths were prescribed to be 20–35 trials long, the block length experienced by the animal was often shorter (8.99 ± 2.78) due to the probabilities changing independently at each spout and the animals switching choices (which begins a new experienced block). We found that when entering a new block (from the animals' perspective), expected uncertainty became lower in the high block relative to the medium block within approximately 5 trials (4.97 ± 1.51). The number of trials the model took to distinguish between reward probabilities in this way was less than the average experienced block lengths (Figure 2-4f,g, 9.03 ± 2.81, $t_{40} = 8.67$, $p < 10^{-10}$, paired t-test). Thus, the updating rate of expected uncertainty allows for the calculation of expected uncertainty and detection of probability changes on timescales relevant to the task and behavior.

We also found evidence of meta-learning in the intra-trial lick behavior. Following no reward, mice consistently licked the chosen spout several times. We found that the number of licks was better explained by unexpected uncertainty from the meta-learning model than by RPE magnitude from the static learning model (Figure 2-3e). In other words, mice licked more when the no reward outcome was most unexpected.

2.3 Discussion

To behave flexibly in dynamic environments, learning rates should vary according to the statistics of those environments (Dayan et al., 2000; Doya, 2002; Kakade and Dayan, 2002; Yu and Dayan, 2005; Behrens et al., 2007). Our model captures differences in learning by estimating expected uncertainty: a moving average of unsigned prediction errors that tracks

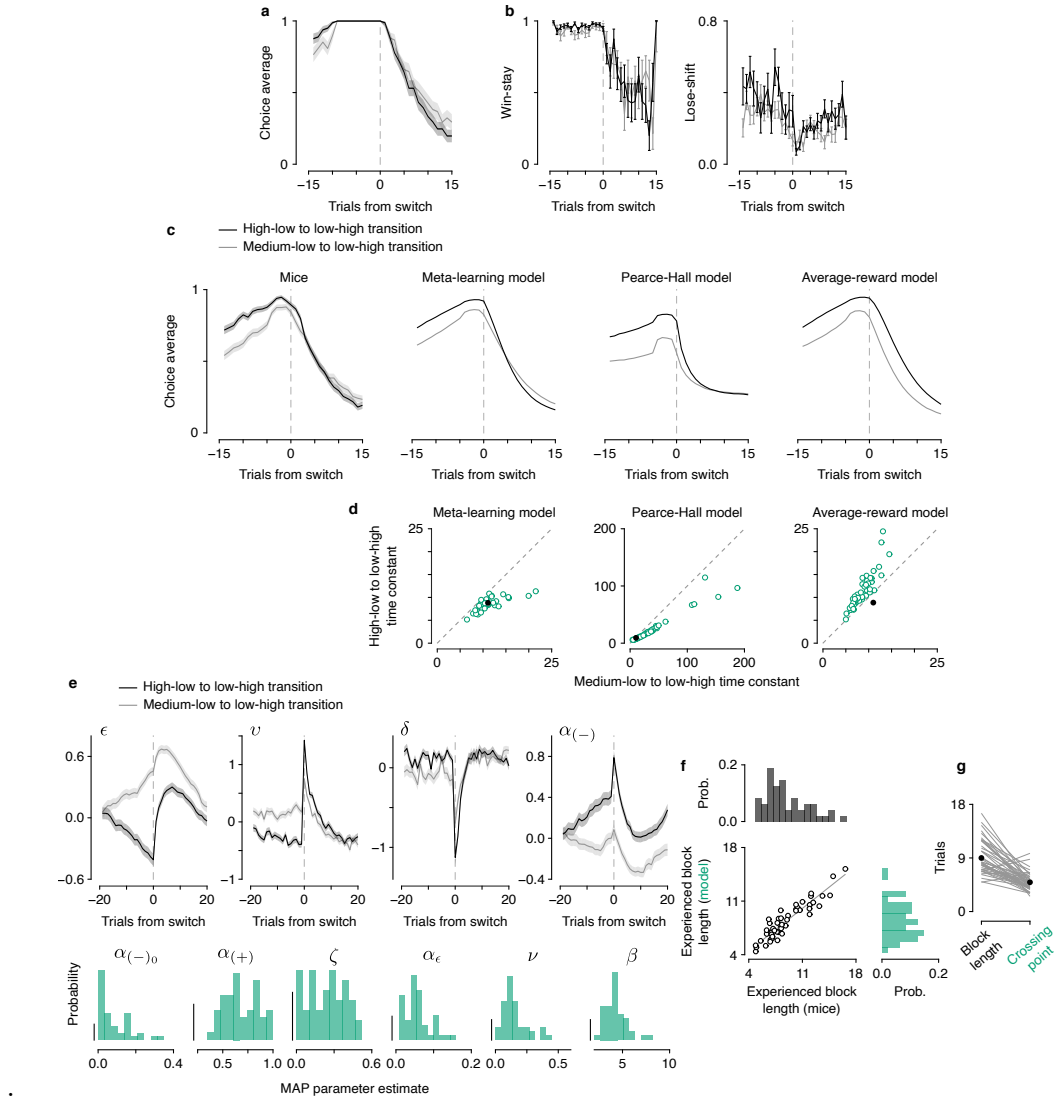

Figure 2-4: **Meta-learning model: data and model comparisons.** (a) Mice showed variable learning rates even when choice history leading up to transitions was identical. (b) Probability of repeating a rewarded choice ("win-stay") and switching following an unrewarded choice ("lose-shift") around transitions. (c) Choice averages (relative to initially-higher spout) for mice and our meta-learning model (both reproduced from Figure 2-3b), compared to two other models with variable learning rates: Pearce-Hall and an average-reward model. (d) The meta-learning model (green) captures transition behavior (black; panel reproduced from Figure 2-1e), whereas the other two models with variable learning rates do not. (e) Top: trial-by-trial dynamics of expected uncertainty (ϵ), unexpected uncertainty (υ), reward prediction error (δ), and negative learning rate ($\alpha_{(-)}$) around transitions in reward probabilities (cf. Figure 2-3b). Mean $\pm$ S.E.M. z-scored values are plotted for each variable. Bottom: maximum *a posteriori* (MAP) parameter estimates. Scale bars: 0.1. (f) Experienced block lengths were similar between mice and models. (g) The number of trials it took the model to discriminate the expected uncertainty in high compared to medium blocks ("crossing point") was less than the experienced block length. Ignores 8 mice that did not distinguish within 30 trials or distinguished before block beginning.

variability in the outcomes of actions. This quantity is used to modulate learning rate by determining how unexpected an outcome is relative to that expected uncertainty. When outcomes are probabilistic but stable, expected uncertainty also slows learning. The model captured observed changes in learning rates that could not be reproduced with an RL model that uses static learning rates.

Several conceptualizations of expected uncertainty have been proposed with different consequences for learning and exploratory behavior (Soltani and Izquierdo, 2019). For example, there can be uncertainty about a specific causal relationship between events in the environment, or between a specific action and the environment. There is evidence that the activity of norepinephrine and acetylcholine neurons may be related to these types of uncertainty (Yu and Dayan, 2005; Hangya et al., 2015; Zhang et al., 2019). It should be noted that both norepinephrine neurons in the locus coeruleus (Szabo and Blier, 2001) and acetylcholine neurons in the basal forebrain (Bengtson et al., 2004) receive functional input from dorsal raphe serotonin neurons.

Here, we studied a more general form of expected uncertainty that tracks variability in outcomes regardless of the specific action taken. This type of uncertainty may apply to learned rules or separately, states in a model-based framework (Bach and Dolan, 2012). It may also be conceptually related to the level of commitment to a belief, which can scale learning in models that learn by minimizing surprise (Payzan-LeNestour and Bossaerts, 2011; Payzan-LeNestour et al., 2013; Faraji et al., 2018). In these ways, our model may approximate inference or change detection in certain behavioral contexts. Our notion of expected uncertainty is also related to reward variance, risk, or outcome uncertainty (Preuschoff et al., 2006; Preuschoff and Bossaerts, 2007; Bach and Dolan, 2012; Monosov, 2020), but with respect to an entire behavioral policy as opposed to a specific action.

Unexpected uncertainty has also been previously defined in numerous ways. In our model, the negative RPE learning rate is a function of recent deviations from expected uncertainty and thus may be most related to a subjective estimate of environmental volatility. This

interpretation is consistent with learning rates increasing as a function of increasing volatility (Behrens et al., 2007). An estimate of volatility may also reflect the surprise that results from the violation of a belief (Payzan-LeNestour and Bossaerts, 2011; Payzan-LeNestour et al., 2013; Faraji et al., 2018). Our observation that brief changes in serotonin neuron firing rates at the time of outcome correlated with unexpected uncertainty is also consistent with previous work showing that serotonin neuron activity correlated with "surprise" when cue-outcome relationships were violated (Matias et al., 2017).

In the meta-learning RL model as we have formulated it, only the negative RPE learning rate is subject to meta-learning. This is an empirical finding and one that may be a consequence of the structure of the task. For example, the reward statistics might result in a saturation of learning from rewards such that its modulation is unnecessary. Asymmetries in the task structure (the absence of trials in which $P(R) = 0.1$ for both spouts) and mouse preference (mice regularly exploited the $P(R) = 0.5$ spout) also result in rewards carrying more information about which spout is "good enough" ($P(R) = 0.9$ or $P(R) = 0.5$). Another possibility, not mutually exclusive with the first, is that learning about rewards and lack thereof could be asymmetric. This asymmetry could result from ambiguity in the non-occurrence of the expected outcome, differences in the magnitude of values of each outcome, or separate learning mechanisms entirely. Similarly, because outcomes are binary in our tasks, learning from negative and positive RPEs could be asymmetric. Alternatively, as described above, this parameterization might just better approximate a more complex cognitive process (e.g., inference) in this specific behavioral context.

Learning is dynamic. Flexible decision making requires using recent experience to adjust learning rates adaptively. The observed foraging behavior demonstrates that learning is not a static process, but a dynamic one. The meta-learning RL model provides a potential mechanism by which recent experience modulates learning adaptively.

2.4 Methods

Animals and surgery. We used 57 male and female mice, backcrossed with C57BL/6J and heterozygous for Cre recombinase under the control of the serotonin transporter gene (Slc6a4$^{tm1(cre)Xz}$, The Jackson Laboratory, 014554; Zhuang et al., 2005). 48 mice (21 female, 27 male) were used for analyzing behavior in the dynamic foraging task, 9 male mice were used analysis of behavior in the dynamic Pavlovian task. Surgery was performed on mice between the ages of 4–8 weeks, under isoflurane anesthesia (1.0–1.5% in O_2) and in aseptic condtions. During all surgeries, custom-made titanium headplates were surgically attached to the skull using dental adhesive (C&B-Metabond, Parkell). After the surgeries, analgesia (ketoprofen, 5 mg kg^{-1} and buprenorphine, 0.05–0.1 mg kg^{-1}) was administered to minimize pain and aid recovery.

For all experiments, mice were given at least one week to recover prior to water restriction. During water restriction, mice had free access to food and were monitored daily in order to maintain 80% of their baseline body weight. All mice were housed in reverse light cycle (12h dark/12h light, dark from 08:00–20:00) and all experiments were conducted during the dark cycle between 10:00 and 18:00. All surgical and experimental procedures were in accordance with the *National Institutes of Health Guide for the Care and Use of Laboratory Animals* and approved by the Johns Hopkins University Animal Care and Use Committee.

Behavioral task. Before training on the dynamic foraging task, water-restricted mice were habituated to head fixation for 1–3 d with free access to water from the provided spouts (two 21 ga stainless steel tubes separated by 4 mm) placed in front of the 38.1 mm acrylic tube in which the mice rested. The spouts were mounted on a micromanipulator (DT12XYZ, Thorlabs) with a custom digital rotary encoder system to reliably determine the position of the lick spouts in XYZ space with 5–10 μm resolution (Bari et al., 2019). Each spout was attached to a solenoid (ROB-11015, Sparkfun) to enable retraction (see Behavioral tasks: dynamic foraging). The odors used for the cues (p-cymene and (−)-carvone) were dissolved

in mineral oil at 1:10 dilution (30μl) and absorbed in filter paper housed in syringe adapters (Whatman, 2.7μm pore size). The adapters were connected to a custom-made olfactometer (Cohen et al., 2012) that diluted odorized air with filtered air by 1:10 to produce a 1.0 L min^{-1} flow rate. The same flow rate was maintained outside of the cue period so that flow rate was constant throughout the task.

Licks were detected by charging a capacitor (MPR121QR2, Freescale) or using a custom circuit (Janelia Research Campus 2019-053). Task events were controlled and recorded using custom code (Arduino) written for a microcontroller (ATmega16U2 or ATmega328). Water rewards were 2–4μl, adjusted for each mouse to maximize the number of trials completed per session and to keep sessions around 60 minutes. Solenoids (LHDA1233115H, The Lee Co) were calibrated to release the desired volume of water and were mounted on the outside of the dark, sound-attenuated chamber used for behavior tasks. White noise (2–60 kHz, Sweetwater Lynx L22 sound card, Rotel RB-930AX two-channel power amplifier, and Pettersson L60 Ultrasound Speaker), was played inside the chamber to block any ambient noise.

During the 1–3 days of habituation, mice were trained to lick both spouts to receive water. Water delivery was contingent upon a lick to the correct spout at any time. Reward probabilities were chosen from the set $\{0, 1\}$ and reversed every 20 trials.

In the second stage of training (5–12 d), the trial structure with odor presentation was introduced. Each trial began with the 0.5 s delivery of either an odor "go cue" ($P = 0.95$) or an odor "no-go cue" ($P = 0.05$). Following the go cue, mice could lick either the left or the right spout. If a lick was made during a 1.5 s response window, reward was delivered probabilistically from the chosen spout. The unchosen spout was retracted at the time of the tongue contacting the other spout so that mice would not try to sample both spouts within a trial. The unchosen spout was replaced 2.5 s after cue onset. Following a no-go cue, any lick responses were neither rewarded nor punished. Reward probabilities during this stage were chosen from the set $\{0, 1\}$ and reversed every 20–35 trials. During this period of training only, water was occasionally manually delivered to encourage learning of the response window and

appropriate switching behavior. Reward probabilities were then changed to $\{0.1, 0.9\}$ for 1–2 days of training prior to introducing the final stage of the task. Rewards were never "baited," as in previous versions of the task (Sugrue et al., 2004; Lau and Glimcher, 2005; Tsutsui et al., 2016; Bari et al., 2019). We did not penalize switching with a "changeover delay." If a directional lick bias was observed in one session, the lick spouts were moved horizontally 50–300μm in the opposite direction prior to the following session.

After the 1.5 s response window, inter-trial intervals were generated as draws from an exponential distribution with a rate parameter of 0.3 and a maximum of 30 s. This distribution results in a flat hazard rate for inter-trial intervals such that the probability of the next trial did not increase over the duration of the inter-trial interval (Luce, 1986). Inter-trial intervals (go-cue on to go-cue on) were 7.45 s on average (range 2.5–32.5 s). As in previous studies, mice made a leftward or rightward choice in greater than 99% of trials (Bari et al., 2019). Mice completed 280 ± 66.6 trials per session (range 79–655 trials).

In the final stage of the task, the reward probabilities assigned to each lick spout were drawn pseudorandomly from the set $\{0.1, 0.5, 0.9\}$ in all the mice from the behavior experiments ($n = 48$), all the mice from the DREADDs experiments ($n = 7$), and half of the mice from the electrophysiology experiments ($n = 2$). The other half of mice from the electrophysiology experiments ($n = 2$) were run on a version of the task with probabilities drawn from the set $\{0.1, 0.4, 0.7\}$. The probabilities were assigned to each spout individually with block lengths drawn from a uniform distribution of 20–35 trials. To stagger the blocks of probability assignment for each spout, the block length for one spout in the first block of each session was drawn from a uniform distribution of 6–21 trials. For each spout, probability assignments could not be repeated across consecutive blocks. To maintain task engagement, reward probabilities of 0.1 could not be simultaneously assigned to both spouts. If one spout was assigned a reward probability greater than or equal to the reward probability of the other spout for 3 consecutive blocks, the probability of that spout was set to 0.1 to encourage switching behavior and limit the creation of a direction bias. If a mouse perseverated on a

spout with reward probability of 0.1 for 4 consecutive trials, 4 trials were added to the length of both blocks. This procedure was implemented to keep mice from choosing one spout until the reward probability became high again.

To minimize spontaneous licking, we enforced a 1 s no-lick window prior to odor delivery. Licks within this window were punished with a new randomly-generated inter-trial interval, followed by a 2.5 s no-lick window. Implementing this window significantly reduced spontaneous licking throughout the entirety of behavioral experiments.

Data analysis. All analyses were performed with MATLAB (Mathworks). All data are presented as mean ± S.D. unless reported otherwise. All statistical tests were two-sided. In Figure 2-1d, the probability that the time constants from the actual behavior belonged to the distribution of simulated behavior time constants was calculated by finding the Mahalanobis distance of the former from the latter, calculating the cumulative density function of the chi-square distribution at that distance, and subtracting it from 1. For all analyses, no-go (dynamic foraging) and CS− (dynamic Pavlovian) cues were ignored and treated as part of the inter-trial interval.

Data analysis: descriptive models of behavior. We fit logistic regression models to predict choice as a function of outcome history for each mouse using the model

$$\log\left(\frac{P(c_r(t))}{1-P(c_r(t))}\right) = \sum_{i=1}^{10} \beta_i^R \left(R_r(t-i) - R_l(t-i)\right) + \sum_{i=1}^{10} \beta_i^{NR} \left(N_r(t-i) - N_l(t-i)\right) + \beta_0,$$

where $c_r(t) = 1$ for a right choice and 0 for a left choice, $R = 1$ for a rewarded choice and 0 for an unrewarded choice, and $N = 1$ for an unrewarded choice and 0 for a rewarded choice. To predict response times (RT), we first z-scored the lick latencies by spout, to correct for differences due to relative spout placement and bias. Then, for each animal we fit the model

$$\mathrm{RT(t)} = \sum_{i=1}^{10} \beta_i^R \left(R_r(t-i) + R_l(t-i)\right) + \beta_0,$$

including a variable for trial number. We fit exponentials with the equation $ae^{-\beta^R_{1:10}/\tau}$ to the regression coefficients, averaged across animals, from the choice and response time models.

Data analysis: generative model of behavior with static learning. We applied a generative RL model of behavior in the foraging task with static learning rates (Daw et al., 2006; Bari et al., 2019). This RL model estimates action values ($Q_l(t)$ and $Q_r(t)$) on each trial to generate choices. Choices are described by a random variable, $c(t)$, corresponding to left or right choice, $c(t) \in \{l, r\}$. The value of a choice is updated as a function of the RPE, and the rate at which this learning occurs is controlled by the learning rate parameter α. Because we observed asymmetric learning from rewards and no rewards (Figure 2-3b), consistent with previous reports (Bari et al., 2019), we included separate learning rates for the different outcomes. For example, if the left spout was chosen, then

$$\begin{aligned} Q_l(t+1) &= \begin{cases} Q_l(t) + \alpha_{(+)}\delta(t), & \text{if } \delta(t) > 0 \\ Q_l(t) + \alpha_{(-)}\delta(t), & \text{if } \delta(t) < 0 \end{cases} \\ Q_r(t+1) &= \zeta Q_r(t), \end{aligned}$$

where $\delta(t) = R(t) - Q_l(t)$ and ζ represents the forgetting rate parameter. The forgetting rate captures the increasing uncertainty about the value of the unchosen spout.

The Q-values are used to generate choice probabilities through a softmax decision function (Daw et al., 2006):

$$\begin{aligned} P(c(t) = r) &= \frac{1}{1 + e^{-\beta(Q_r(t) - Q_l(t) + bias)}}, \\ P(c(t) = l) &= 1 - P(c(t) = r), \end{aligned}$$

where β, the "inverse temperature" parameter, controls the steepness of the sigmoidal function. In other words, β controls the level of exploration versus exploitation with respect to the relative action values.

Data analysis: generative model of behavior with meta-learning. We observed mouse behavior that the static learning model failed to capture and that suggested that learning rate was not constant over time. Thus, we added a component to the model that modulates RPE magnitude and $\alpha_{(-)}$ ("meta-learning"). Because learning should be slow in stable but variable environments, expected uncertainty scaled RPEs, such that learning is decreased when expected uncertainty is high. If the left spout was chosen, the values of actions were updated according to

$$\begin{aligned} Q_l(t+1) &= \begin{cases} Q_l(t) + \alpha_{(+)}\delta(t)(1-\epsilon(t)), & \text{if } \delta(t) > 0 \\ Q_l(t) + \alpha_{(-)}(t)\delta(t)(1-\epsilon(t)), & \text{if } \delta(t) < 0 \end{cases} \\ Q_r(t+1) &= \zeta Q_r(t), \end{aligned}$$

where ϵ is an evolving estimate of expected uncertainty calculated from the history of unsigned RPEs:

$$\begin{aligned} \upsilon(t) &= |\delta(t)| - \epsilon(t), \\ \epsilon(t+1) &= \epsilon(t) + \alpha_\upsilon \upsilon(t). \end{aligned}$$

The rate of RPE magnitude integration is controlled by α_υ. Deviations from the expected uncertainty are captured by unexpected uncertainty, υ, and may indicate that a change has occurred in the environment. Changes in the environment should drive learning to adapt behavior to new contingencies so $\alpha_{(-)}$ varies as a function of how surprising recent outcomes are:

$$\alpha_{(-)}(t) = \begin{cases} \alpha_{(-)}(t-1) & \text{if } \delta(t) > 0 \\ \psi(\upsilon(t) + \alpha_{(-)_0}) + (1-\psi)(\alpha_{(-)}(t-1)) & \text{if } \delta(t) < 0 \end{cases}$$

where $\alpha_{(-)_0}$ is the baseline learning rate from no reward and ψ controls how quickly unexpected uncertainty is integrated to update $\alpha_{(-)}$. As it is formulated, $\alpha_{(-)}$ increases after surprising no reward outcomes. This learning rate was not allowed to be less than 0, such that

$$\alpha_{(-)}(t) = 0, \text{if } \alpha_{(-)}(t) < 0$$

To generate choice probabilities, the Q-values were fed into the same softmax decision function as the static-learning model.

We also examined two other meta-learning models from the Q-learning family of RL models. The first is an updated form of the opponency model (Daw et al., 2002) referred to as the global reward state model (Wittmann et al., 2020). In this model, a global reward history variable influences learning from rewards and no rewards asymmetrically, as those outcomes carry different amounts of information depending on the richness of the environment. In this model, the value of a chosen action, for example Q_l, is updated according to

$$Q_l(t+1) = Q_l(t) + \alpha\delta(t),$$
$$Q_r(t+1) = \zeta Q_r(t),$$

while the unchosen action value, Q_r is forgotten with rate ζ. The prediction error, δ, is calculated by

$$\delta(t) = R(t) - Q_l(t) + \omega\bar{R}(t),$$

where R is the outcome, $\bar{R}$ is a global reward history term and ω is a weighting parameter that can be positive or negative. $\bar{R}$ is updated on each trial:

$$\bar{R}(t+1) = \bar{R}(t) + \alpha_{\bar{R}}(\bar{R}(t) - R(t)).$$

Here, $\alpha_{\bar{R}}$ is the learning rate for the global reward term. The learned action values are converted into choice probabilities using the same softmax decision function described above.

The second model we tested is an adapted Pearce-Hall model (Pearce and Hall, 1980) in which the learning rate is a function of RPE magnitude. If the left action is chosen, Q_l is updated by the learning rule

$$\begin{aligned} Q_l(t+1) &= \begin{cases} Q_l(t) + \kappa_{(+)}\alpha(t)\delta(t), & \text{if } \delta(t) > 0 \\ Q_l(t) + \kappa_{(-)}\alpha(t)\delta(t), & \text{if } \delta(t) < 0 \end{cases} \\ Q_r(t+1) &= \zeta Q_r(t), \end{aligned}$$

where $\kappa_{(+)}$ and $\kappa_{(-)}$ are the salience parameters for rewards and no rewards, respectively. Having separate salience parameters is a modification of the original model that we made to improve fit and mirror the asymmetry in our own meta-learning model and the global reward state model. The learning rate α is updated a function of RPE:

$$\alpha(t+1) = \alpha(t) + \eta(\alpha(t) - |\delta(t)|).$$

Here, η controls the rate at which the learning rate is updated. In this way, the model enhances learning rates when the recent average of RPE magnitudes is large. This approach contrasts with our meta-learning model which diminishes the learning rate as a result of large recent RPE magnitudes if they are consistent.

Data analysis: Model fitting. We fit and assessed models using MATLAB (Mathworks) and the probabilistic programming language, Stan (https://mc-stan.org/) with the MATLAB

interface, MatlabStan (https://mc-stan.org/users/interfaces/matlab-stan). Stan was used to construct hierarchical models with mouse-level hyperparameters to govern session-level parameters. For each session, each parameter in the model (for example, α_ϵ for the meta-learning model) was modeled as a draw from a mouse-level distribution with mean μ and variance σ. Models were fit using noninformative (uniform distribution) priors for session-level parameters ($[0, 1]$ for all parameters except β which was $[0, 10]$) and weakly informative ($\mu \sim \mathcal{N}(0, 1)$, $\sigma \sim$ half-Cauchy$(0, 3)$) priors for mouse-level hyperparameters. For some mice with fewer sessions, more informative mouse-level hyperparmeters were used to achieve model convergence under the assumption that individual mice behave similarly across days. This hierarchical construction mitigated the typical variability of point estimates for session-level parameters that results from other methods of estimation. Stan uses full Bayesian statistical inference to generate posterior distributions of parameter estimates using Hamiltonian Markov chain Monte Carlo sampling (Carpenter et al., 2017). The parameters for updating expected uncertainty, α_υ, and for updating the negative RPE learning rate, ψ, were ordered such that $\psi > \alpha_\upsilon$. The ordering operated under the assumption that learning rate should be integrated more quickly to detect change. The ordering also helped models to converge more quickly.

Data analysis: extracting model parameters and variables, behavior simulation. For extracting model variables (like expected uncertainty), we took at least 4,000 draws from the Hamiltonian Markov Chain Monte Carlo samples of session-level parameters, ran the model agent through the task with the actual choices and outcomes, and averaged each model variable across runs. For comparisons of individual parameters across behavioral and neural models, we obtained maximum *a posteriori* parameter values by approximating the mode of the distribution: binning the values in 50 bins and taking the median value of the most populated bin. For simulations of behavior, we took at least 4,000 draws from the Hamiltonian Markov Chain Monte Carlo samples of mouse-level parameters and simulated behavior and outcomes in a number of random sessions per sample. For the transition analysis, that number was proportional to the number of rare transitions that each animal

contributed to the actual data. For other analyses that number was fixed.

Data analysis: linear mixed effect models. To analyze the changes in transition behavior we constructed a linear mixed effects model that predicted choice averages after transition points as a function of trial since transition, transition type, and the interaction between the two. Choice probabilities were first z-scored in order to center the data. The model is described by the following Wilkinson notation:

$$choice\,averages \sim trial\,from\,transition * transition\,type.$$

Chapter 3

Dorsal raphe serotonin neurons track uncertainty to modulate learning rate

Abstract

Learning rates are variable and, normatively, should be guided by uncertainty. We previously demonstrated that mice learn at variable rates as a function of recent reward statistics. The modulation of learning rates was effectively modeled as a function of two types of uncertainty. Expected uncertainty stabilized behavior by mitigating learning in the face of probabilistic outcomes. Unexpected uncertainty permitted detection of change in the contingencies between action and outcomes, enhancing learning accordingly. Earlier work suggests that serotonin neurons are involved in modulating learning rates. Here, we show that serotonin neuron activity tracks both types of uncertainty from our meta-learning model in order to drive learning rates in mice behaving in a dynamic foraging task.

3.1 Introduction

Several theories propose that neuromodulatory systems enable meta-learning (Daw et al., 2002; Doya, 2002; Yu and Dayan, 2005). One such system comprises a small number of serotonin-releasing neurons (on the order of 10^4 in mice; Ishimura et al., 1988) with extensive axonal projections. This small group of cells affects large numbers of neurons in distributed

regions (Steinbusch, 1981; Jacobs and Azmitia, 1992; Ren et al., 2018; Awasthi et al., 2020) that are responsible for learning and decision making. The activity of these neurons changes on behaviorally-relevant timescales—both fast (hundreds of milliseconds) and slow (tens of seconds; Liu et al., 2014; Cohen et al., 2015; Li et al., 2016; Matias et al., 2017). Serotonin receptor activation can induce short-term changes in excitability (Andrade and Haj-Dahmane, 2020) as well as long-lasting synaptic plasticity (Lesch and Waider, 2012).

Prior research demonstrates that serotonin neurons modulate flexible behavior in changing environments (Clarke et al., 2004, 2007; Boulougouris and Robbins, 2010; Bari et al., 2010; Brigman et al., 2010; Matias et al., 2017; Iigaya et al., 2018). Serotonin axon lesions (Clarke et al., 2004, 2007) or reversible inactivation of dorsal raphe serotonin neurons (Matias et al., 2017) impaired behavioral adaptation to changes in action- or stimulus-outcome mappings. Importantly, in these experiments, animals were still capable of adapting their behavior, but did so more slowly. Conversely, brief excitations of serotonin neurons in a probabilistic choice task enhanced learning rates after long intervals between outcomes (Iigaya et al., 2018). These studies show that serotonin neurons modulate how quickly an animal adapts to a change in causal relationships in the environment. Thus, serotonin neurons may guide learning using the statistics of recent outcomes. More specifically, they may track the expected and unexpected uncertainty about a behavioral policy in the manner proposed by our meta-learning model. However, a mechanistic understanding of the relationship between serotonin neuron activity and uncertainty-driven meta-learning has not been established.

We recorded action potentials from dorsal raphe serotonin neurons of mice behaving in our dynamic foraging task. We found that the activity of approximately half of serotonin neurons correlated with the "expected uncertainty" variable from our meta-learning model on long timescales (tens of seconds to minutes) and "unexpected uncertainty" at the time of outcome. We also found these relationships were maintained in a different behavioral context using a Pavlovian version of the task. Simulated removal of meta-learning from the model predicted specific changes in learning that were reproduced by chemogenetic inhibition of

dorsal raphe serotonin neurons during dynamic foraging. Thus, we demonstrate a quantitative link between serotonin neuron activity and behavior.

3.2 Results

Serotonin neuron firing rates correlate with expected uncertainty

To quantify the link between serotonin neurons and meta-learning, we recorded action potentials from dorsal raphe serotonin neurons in mice performing the foraging task (66 neurons from 4 mice). To identify serotonin neurons, we expressed the light-gated ion channel, channelrhodopsin-2, under the control of the serotonin transporter promoter in *Slc6a4*-Cre (also known as *Sert*-Cre) mice (Figures 3-1a, 3-2a). We delivered light stimuli to the dorsal raphe to "tag" serotonin neurons at the end of each recording (Figures 3-1b, 3-2b,c). We calculated firing rates during the inter-trial intervals and compared the activity to the behavioral model variables. We found a significant relationship between firing rate and expected uncertainty in 53% (35 of 66) of serotonin neurons (Figure 3-1c-e; regression of inter-trial interval firing rates on expected uncertainty). When we regressed out slow, monotonic changes in firing rates and expected uncertainty over the course of the session, this relationship held (Figure 3-2d). In contrast, we did not find such prevalent relationships in a multivariate regression of firing rates on other latent model variables, such as relative value or RPE (Figure 3-2e).

Remarkably, firing rates were stable within inter-trial intervals. Dividing expected uncertainty into terciles, we found that serotonin neuron firing rates were relatively constant as time elapsed within inter-trial intervals (Figure 3-1f; regression coefficient $= 1.9 \times 10^{-6}$ from a linear model of tercile difference on time in inter-trial interval). Because expected uncertainty evolved somewhat slowly as a function of RPE magnitude and the activity of neurons on this timescale (tens of seconds) fluctuated slowly as well, the two may be similarly autocorrelated (Elber-Dorozko and Loewenstein, 2018). To control for spurious correlations

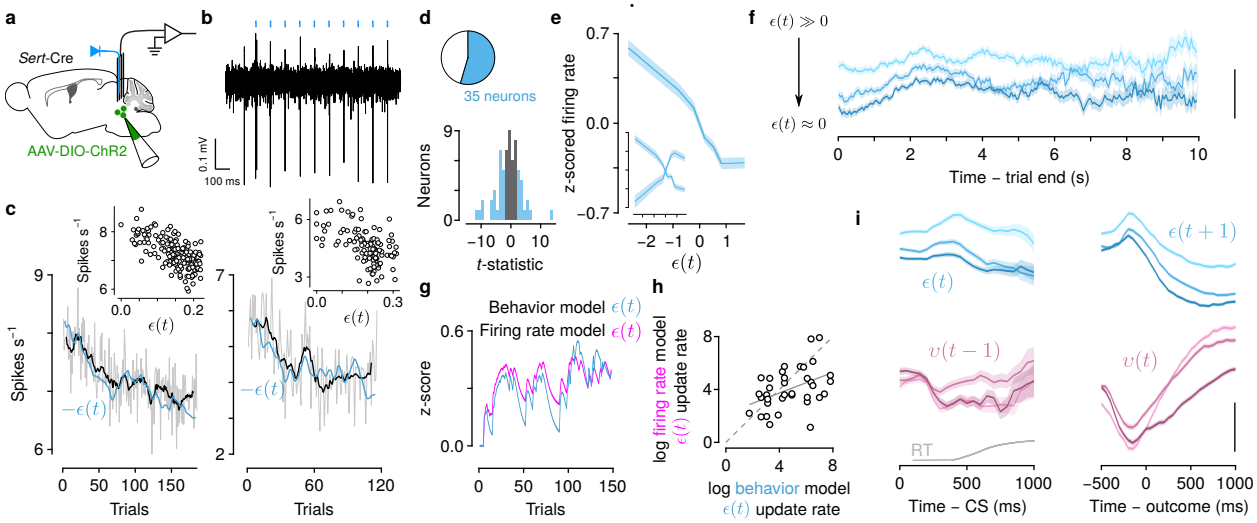

Figure 3-1: **Serotonin neuron firing rates correlate with expected uncertainty on slow timescales and unexpected uncertainty on fast timescales.**(a) Schematic of electrophysiological recording of identified serotonin neurons. (b) Example "tagging" of a serotonin neuron, using channelrhodopsin-2 stimulation. (c) Two example neurons showing negative correlations between inter-trial interval firing rates and expected uncertainty ($-\epsilon(t)$ is plotted). Insets show trial-by-trial relationships. (d) t-statistics across all neurons from a linear regression, modeling firing rates as a function of $\epsilon(t)$. Blue bars and slice indicate neurons with significant regression coefficients. (e) Population z-scored firing rates varied with $\epsilon(t)$. Inset shows population split by positive and negative correlations. "Sign-flipped" plot, combining across these neurons, was used for analysis in (f). (f) z-scored firing rates of serotonin neurons split by $\epsilon(t)$ tercile. Scale bar: 0.5 z-score. (g) Example dynamics of $\epsilon(t)$ estimated from behavior and neuronal firing rates. (h) Log-log plot of the expected uncertainty update rate (ψ) from each neuron's firing rate model and the behavioral model derived from simultaneous choice behavior. (i) Within-trial dynamics of expected ($\epsilon(t)$, $\epsilon(t+1)$, top row) and unexpected ($\upsilon(t-1)$, $\upsilon(t)$, bottom row) uncertainty, aligned to go cue (CS, left column) and outcome (right column). Scale bar: 0.5 z-score. Gray curve: response time (RT) distribution (cut off at 1 s).

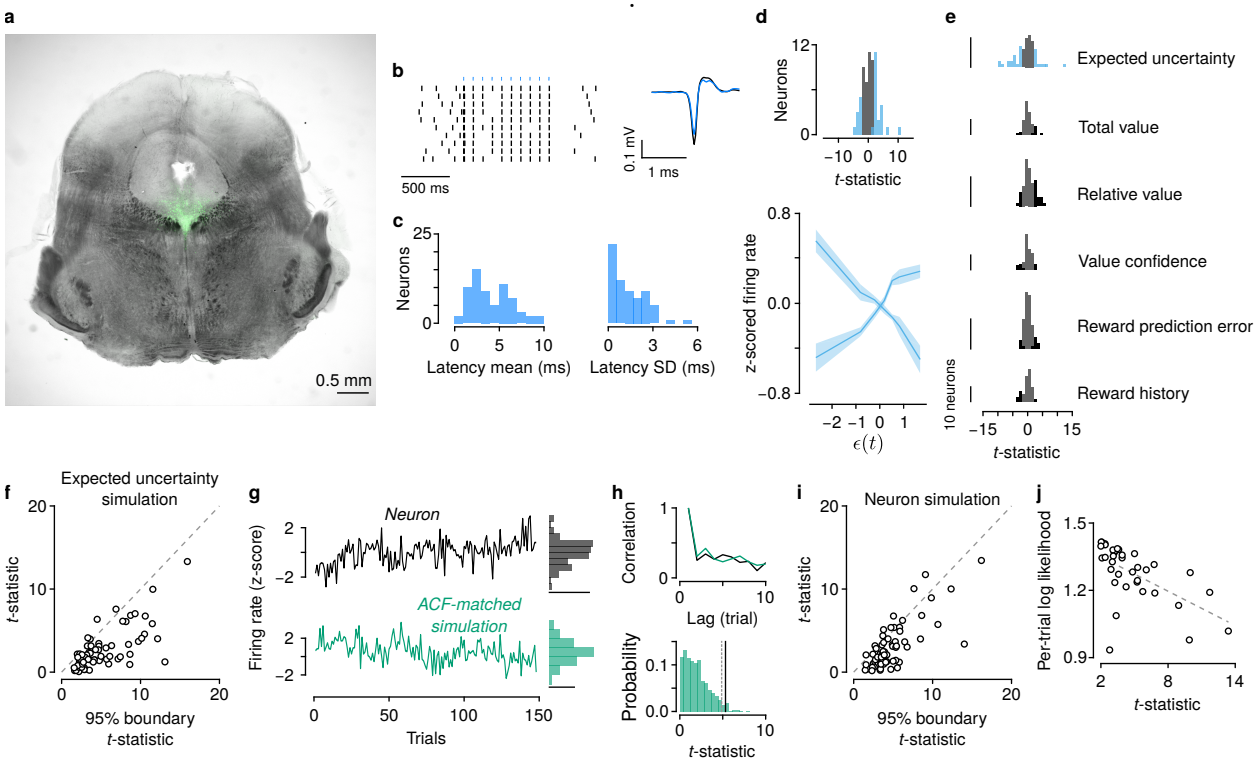

Figure 3-2: **Serotonin neuron firing rates correlate with expected uncertainty.** (a) Representative histological section of the midbrain from electrophysiological experiments, showing ChR2-EYFP expression (green) in dorsal raphe serotonin neurons. (b) Example of identified serotonin neuron firing in response to most light stimuli activating ChR2 with short latency and similar extracellular action potential waveform. (c) Mean and SD of firing latency of identified serotonin neurons. (d) Regression results as in Figure 3-1d,e, removing slow, session-long trends. (e) Distributions of t-statistics of regressors in a multivariate generalized linear model of inter-trial interval firing rate. (f) t-statistics from neurons compared with true and simulated expected uncertainty. (g) Example simulated neuron with an autocorrelation function (ACF) matched to the real neuron. Probability density scale bars: 0.2. (h) Top: ACF matching between real neuron and simulations. Bottom: distribution of t-statistic from the real neuron (black line) and simulations (green). Dashed gray line shows 95% boundary from the distribution of simulations. (i) t-statistics from real and simulated neurons compared with expected uncertainty. (j) Success of firing rate model fit correlates with t-statistic comparing firing rate to behavior-model-derived expected uncertainty ($R^2 = 0.34$, Spearman's $\rho = -0.65$, $p < 10^4$).

due to comparison of two autocorrelated variables, we first compared the actual neural data to simulated expected uncertainty terms (Figure 3-2f). We found stronger statistical relationships across the population with the actual expected uncertainty than with simulated values. Additionally, we simulated neural activity with quantitatively-matched autocorrelation functions to the real neurons and compared this activity to the actual expected uncertainty values. Again, we found stronger statistical relationships in the real data as opposed to the simulated data (Figure 3-2g-i).

To further examine the robustness of this relationship, we fit the meta-learning model to the inter-trial interval firing rates of neurons that had a significant correlation with expected uncertainty. The meta-learning algorithm was essentially the same as before, but we fit firing rates as a function of expected uncertainty as opposed to fitting choices as a function of relative action values. Here, we found that the updating rate for expected uncertainty from the firing rate model covaried with the same parameter from the choice model across sessions (Figure 3-1g,h). Additionally, how well the model fit to the firing rates was predicted by how well-correlated the firing rates were to the expected uncertainty variable from the behavioral model (Figure 3-2j). This result suggests that the neural and behavior data, independently, predict similar expected uncertainty dynamics.

Serotonin neuron firing rates correlate with unexpected uncertainty at outcomes

How does the presence or absence of reward update the slowly-varying firing rates of serotonin neurons? According to the model, expected uncertainty changes as a function of unexpected uncertainty. In particular, the model thus predicts a firing rate change at the time of outcome that could be used to update expected uncertainty.

To test this, we calculated firing rates of serotonin neurons within trials, while mice made choices and received outcomes. We found that firing rate changes on fast timescales (hundreds of ms) correlated with expected uncertainty ($\epsilon(t)$) throughout the period when

mice received go cues and made choices (Figure 3-1i). These correlations persisted during the outcome (reward or no reward), as $\epsilon(t)$ updated to its next value ($\epsilon(t+1)$). In contrast, firing rates correlated with unexpected uncertainty ($\upsilon(t)$) primarily during the outcome, but not during the go cue ($\upsilon(t-1)$; Figure 3-1i). Thus, brief firing rate changes in serotonin neurons could be integrated to produce more slowly-varying changes. In this computation, firing rates may be interpreted as encoding two forms of uncertainty, one slowly varying (ϵ), one more transient (υ).

Serotonin neuron firing rates correlate with uncertainty in a Pavlovian task

Based on the results from the first experiment, we made two predictions. First, we hypothesized that correlations between serotonin neuron activity and expected uncertainty generalize to other behavioral tasks. To test this, we trained 9 mice on a Pavlovian version of the task in which an odor cue predicted probabilistic reward after a 1-s delay (Figure 3-3a). The probability of reward changed in blocks within each session (Figure 3-3b). This task required no choice to be made. Rather, mice simply licked toward a single water-delivery spout in anticipation of a possible reward.

The number of anticipatory licks during the delay between cue and outcome (presence or absence of reward) reflected recent reward history (Figure 3-3c). To estimate ongoing expected uncertainty in this task, we modified the meta-learning model to generate anticipatory licks as a function of the value of the cue (Figure 3-4a). While the model was capable of explaining behavior, interestingly, we found no clear behavioral evidence of variable learning rates (Figure 3-4b,c). However, recordings of dorsal raphe serotonin neurons from mice behaving in this task revealed that the activity of these neurons correlated with expected uncertainty at similar rates to those recorded in the dynamic foraging task (Figures 3-3d-f, 3-4d-f; 68%, 28 of 41 neurons from 5 mice). Similar to observations in the foraging task, neurons in the Pavlovian task showed stable firing rates within inter-trial intervals (Figure 3-3g; regression

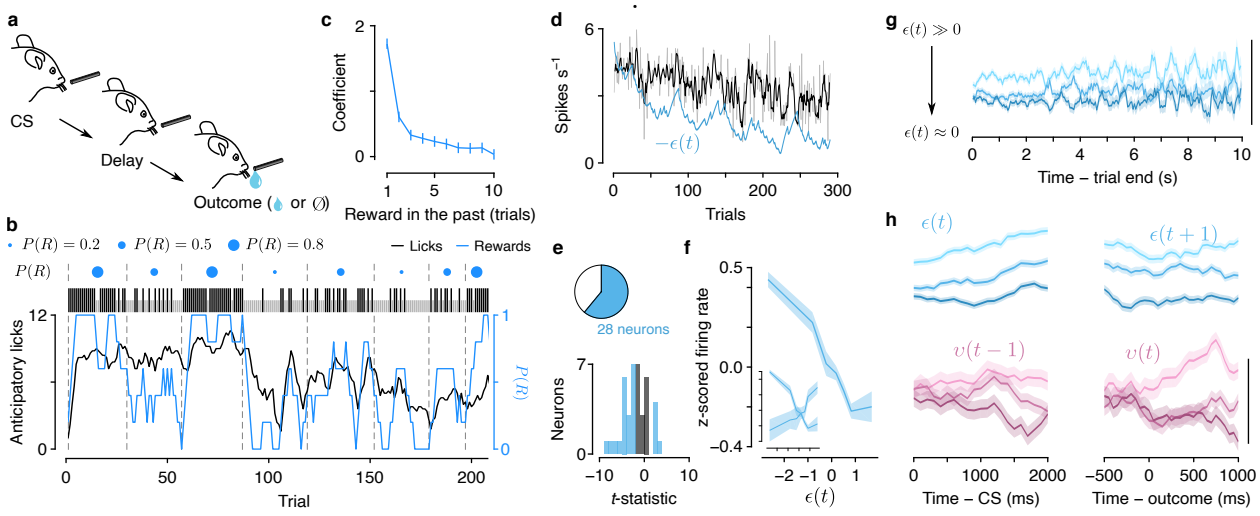

Figure 3-3: **Serotonin neuron firing rates correlate with expected and unexpected uncertainty in a dynamic Pavlovian task.**(a) Schematic of Pavlovian task in which the probability of reward ($P(R)$) varied over trials. (b) Example behavior showing anticipatory licking, in the delay before outcome, as $P(R)$ varied. Black ticks: rewarded trials. Gray ticks: unrewarded trials. (c) Linear regression coefficients of licking rate on reward history. (d) Example serotonin neuron showing a negative correlation between inter-trial interval firing rates and expected uncertainty ($-\epsilon(t)$ is plotted). (e) t-statistic from linear regression, modeling firing rate as a function of $\epsilon(t)$ as in Figure 3-1d. (f) Population "tuning curves," as in Figure 3-1e. (g) Stable firing rates within inter-trial intervals, as in Figure 3-1f. Scale bar: 0.5 z-score. (h) Expected and unexpected uncertainty z-scored firing rates, as in Figure 3-1i. Scale bar: 0.5 z-score.

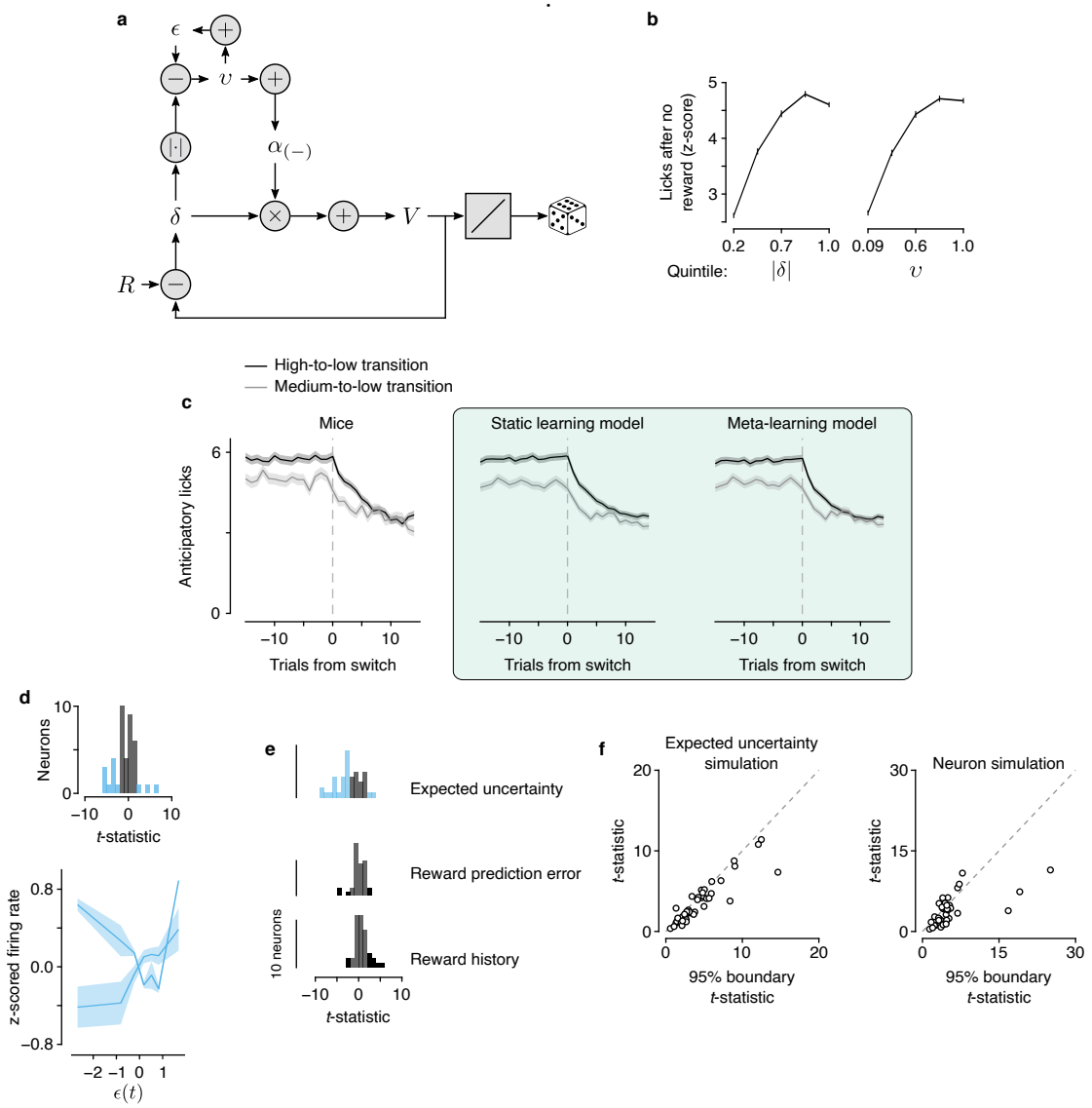

Figure 3-4: **Serotonin neuron firing rates correlate with expected uncertainty in a dynamic Pavlovian task.** (a) Schematic of meta-learning model applied to behavior in the dynamic Pavlovian task. The value (V) of the stimulus is updated analogously to the way action values (Q_l and Q_r) are updated in the dynamic foraging task. V is mapped to licks through a linear scaling and sampling from a Poisson distribution. (b) Lick rate after no reward scales with unsigned RPE ($|\delta|$, regression coefficient $= 0.68$, $R^2 = 0.11$) and unexpected uncertainty (ϵ, regression coefficient $= 0.51$, $R^2 = 0.083$). (c) Transition behavior in the dynamic Pavlovian task when probabilities changed from high to low or medium to low. Left: mice. Middle: static learning model simulated behavior. Right: meta-learning model simulated behavior. (d) Regression results as in Figure 3-3e,f, removing monotonic, session-long trends. (e) Distributions of t-statistics of regressors in a multivariate generalized linear model of inter-trial interval firing rate. (f) Left: t-statistics from neurons compared with true and simulated expected uncertainty. Right: t-statistics from real and simulated neurons compared with expected uncertainty.

coefficient $= -3.1 \times 10^{-6}$ from a linear model of tercile difference on time in inter-trial interval). Serotonin neuron firing rates also correlated with expected uncertainty throughout its update interval, and with unexpected uncertainty at the time of the outcome (Figure 3-3h). Thus, the nervous system may maintain running estimates of two forms of uncertainty that generalizes across behavioral tasks.

Serotonin neuron inhibition disrupts meta-learning

In our second prediction from the dynamic foraging experiment, we asked whether inactivating serotonin neurons rendered mice unable to adjust learning rates. The meta-learning model makes specific predictions about the role of uncertainty in learning. To test the predictions of the model under the hypothesis that serotonin neurons encode expected uncertainty, we expressed an inhibitory designer receptor exclusively activated by designer drugs (DREADD) conjugated to a fluorophore (hM4Di-mCherry) in dorsal raphe serotonin neurons (Figure 3-5a). *Sert*-Cre mice received injections of a Cre-dependent virus containing the receptor (AAV5-hSyn-DIO-hM4D(Gi)-mCherry, $n = 3$ mice) into the dorsal raphe. Control *Sert*-Cre mice were injected with the same virus containing only the fluorophore ($n = 4$ mice). On consecutive days, mice received an injection of vehicle (0.5% DMSO in 0.9% saline), the DREADD ligand agonist 21 (3 mg kg^{-1} in vehicle; Chen et al., 2015; Thompson et al., 2018), or no injection. Because the simultaneous changes of reward probabilities were rare, we modified the task to include them with slightly higher frequency.

To quantify the change in behavior predicted by the model, we first fit the model to mouse behavior on vehicle injection days and used those parameters to simulate behavior. We then simulated a lesion by fixing expected and unexpected uncertainty to 0 (essentially fixing the negative RPE learning rate to its median value) and simulated behavior again (Figure 3-5b). The simulated lesion diminished the differences in transition speed between the pre-transition reward conditions (Figure 3-5c,e).

On days with agonist 21 injections, mice expressing hM4Di in serotonin neurons demon-

strated changes in learning at transitions (Figure 3-5d,f; effects of trial from transition $F_{1,58} = 44.3$, $p < 10^{-7}$, drug condition $F_{1,58} = 41.5$, $p < 10^{-7}$, trial $\times$ transition type interaction $F_{1,58} = 6.15$, $p = 0.016$, and transition type $\times$ drug condition interaction $F_{1,58} = 21.2$, $p < 10^{-4}$, linear mixed effects model) matching the predictions of the simulated lesion model (Figure 3-5c,e; effects of trial from transition $F_{1,186} = 180.4$, $p < 10^{-28}$, drug condition $F_{1,186} = 18.1$, $p < 10^{-4}$, trial $\times$ transition type interaction $F_{1,186} = 7.51$, $p = 0.0067$, and transition type $\times$ drug condition interaction $F_{1,186} = 8.28$, $p = 0.0045$). The same experiment in mice expressing a fluorophore alone in serotonin neurons showed no effect of agonist 21 (Figure 3-5h,j; effect of trial from transition $F_{1,58} = 95.8$, $p < 10^{-13}$ and effect of transition type $F_{1,58} = 5.55$, $p = 0.022$), consistent with simulations from the meta-learning model fit separately to vehicle and agonist 21 behavior (Figure 3-5g,i; effect of trial from transition $F_{1,250} = 483$, $p < 10^{-59}$ and effect of trial type $\times$ transition type interaction $F_{1,58} = 7.02$, $p = 0.0086$). Serotonin neuron inhibition did not slow response times (Figure 3-6b), change how outcomes drove response times (Figure 3-6c), nor cause mice to lick during inter-trial intervals. Thus, the observed effects of reversible inhibition are consistent with a role for serotonin neurons signaling uncertainty.

3.3 Discussion

The activity of the majority of identified serotonin neurons correlated with the expected uncertainty and unexpected uncertainty variables from our meta-learning model when fit to dynamic foraging behavior. These relationships held in a different behavioral context, with similar fractions of serotonin neurons tracking expected uncertainty in a dynamic Pavlovian task. During dynamic foraging, chemogenetic inhibition of serotonin neurons caused changes in choice behavior that were consistent with the changes in learning predicted by removing meta-learning from the model.

We did not find any behavioral evidence in the dynamic Pavlovian behavior that distinguished meta-learning from static learning. It may be that differences in these models are

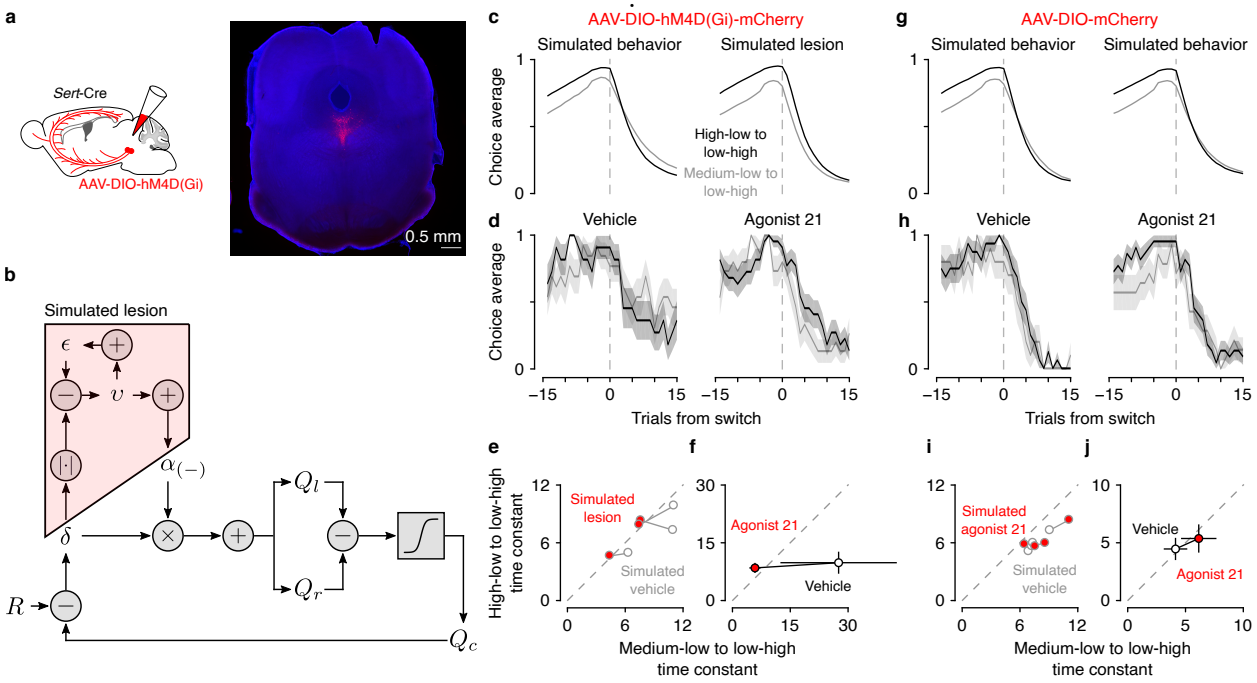

Figure 3-5: **Serotonin neuron inhibition disrupts meta-learning.** (a) Schematic of experiment to reversibly inactivate serotonin neurons and representative expression of hM4Di-mCherry in dorsal raphe serotonin neurons. (b) Schematic of simulated lesion, in which models were fit to mouse behavior, and then meta-learning variables (i.e., ϵ and υ) were set to zero. (c) Simulated behavior with meta-learning intact, fit to vehicle behavior (left) and simulated lesion (right). (d) Mouse behavior with vehicle injections (control experiment) and drug (agonist 21). Lines are mean choice probability and shading is Bernoulli S.E.M. (e) Exponential time constants for transitions from simulated behavior and simulated lesions. (f) Time constants from mice (with 95% C.I.). (g) Simulated behavior from mice expressing mCherry in serotonin neurons with vehicle (left) and agonist 21 (right) injections. (h) Mouse behavior with vehicle injections (control experiment) and drug (agonist 21). (i) Simulation time constants from fluorophore-control mice. (j) Time constants from fluorophore-control mice (with 95% C.I.).

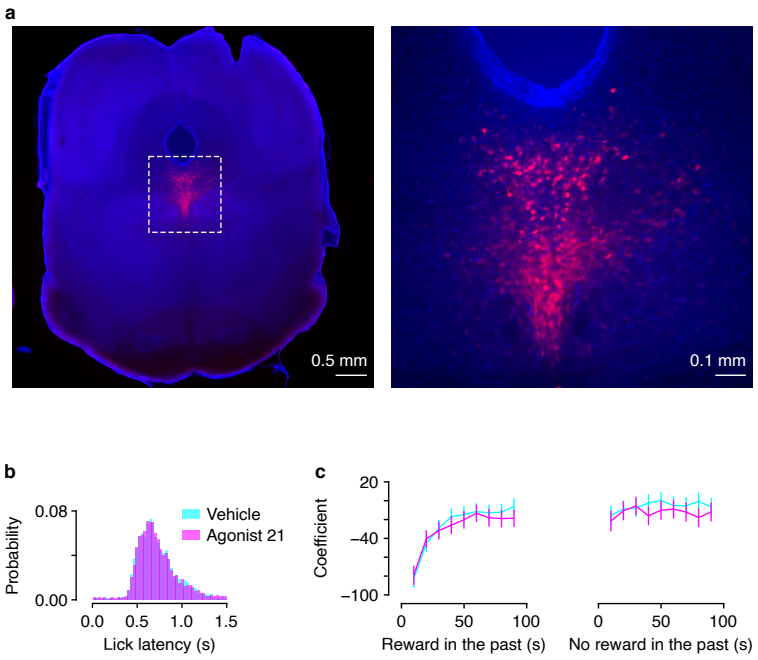

Figure 3-6: **Serotonin neuron inhibition does not affect lick latency.**(a) Representative histological section showing DREADD expression in dorsal raphe serotonin neurons (reproduced from Figure 3-5a). Dashed box in the left image indicates higher-magnification image on the right. (b) No difference in lick latency comparing vehicle injections to agonist 21 injections (paired t-test, $t_2 = -0.39$, $p = 0.73$). (c) Regression coefficients modeling lick latency as a function of reward (left) or no reward (right) in the past, for vehicle or agonist 21 injections.

not observable in this behavior. Also, the dynamic foraging task engages regions of the brain that are not necessary for the dynamic Pavlovian task (Bari et al., 2019). Consequently, uncertainty may be incorporated in other ways to drive behavior. Alternatively, the brain may keep track of statistics of the environment that are not always used in behavior.

While serotonin neuron firing rates change on multiple timescales (Cohen et al., 2015), the observed changes that correlated with expected uncertainty occurred over relatively long periods of time. How rapidly RPE magnitudes are integrated tracks variability in outcomes on a timescale relevant to experienced block lengths. Consequently, deviations from expected uncertainty reliably indicate changes in reward probabilities. In addition to the computational relevance of activity on this timescale, serotonin neuron firing rate changes may be optimized for the nervous system to implement these computational goals. Slow changes in serotonin neuron activity could enable gating or gain control mechanisms (Shimegi et al., 2016; Azimi et al., 2020), bidirectional modulation of relevant inputs and outputs (Avesar and Gulledge, 2012; Stephens et al., 2014, 2018), or other previously-observed circuit mechanisms that modulate how new information is incorporated (Marder, 2012) to drive flexible behavior.

We also observed changes in serotonin neuron activity on shorter timescales that correlated with both expected and unexpected uncertainty. The timing of these brief signals may be important to update the slower dynamics correlated with expected uncertainty, as predicted from the model (i.e., ϵ essentially integrates υ). Alternatively, serotonin neurons could "multiplex" across timescales, whereby brief changes in firing rates may have different downstream functions than slower changes.

Our findings and conceptual framework, including the effect on learning from worse-than-expected outcomes, are consistent with previous observations and manipulations of serotonin neuron activity. In a Pavlovian reversal task, changes in cue-outcome mappings elicited responses from populations of serotonin neurons that decayed as mice adapted their behavior to the new mapping (Matias et al., 2017). Chemogenetic inhibition of serotonin neurons in this task impaired behavioral adaptation to a cue that predicted reward prior to the reversal

but not after. The manipulation did not affect behavior changes in response to the opposite reversal. In reversal tasks in which action-outcome contingencies were switched, lesions or pharmacological manipulations of serotonin neurons also resulted in impairments of adaptive behavior at the time of reversal (Clarke et al., 2004, 2007; Boulougouris and Robbins, 2010; Bari et al., 2010; Brigman et al., 2010). Specifically, lesioned animals continued to make the previously-rewarded action. These findings are consistent with a role for serotonin neuron activity in tracking expected uncertainty and driving learning from worse-than-expected outcomes.

More recent work demonstrated that serotonin neuron activation increased the learning rate after longer intervals between outcomes, but that learning after shorter intervals was already effectively saturated (i.e., win-stay, lose-shift; Iigaya et al., 2018). An intriguing possibility (similarly raised by the authors), is that serotonin neurons mediate the contributions of faster, working-memory-based learning, and slower, plasticity-dependent learning that may map onto model-based and model-free learning, respectively. There is evidence to support the existence of both model-free and model-based systems of learning and decision making in the brain (Balleine and Dickinson, 1998; Daw et al., 2005; Lee et al., 2014; Miller et al., 2017), although their distinction and conceptualizations are debated (Miller et al., 2018). One theory proposed that the arbitration between these two processes is guided by their relative reliability (Lee et al., 2014). Interestingly, the arbitration process relied on comparisons of the absolute value of RPEs from the respective processes. The possibility arises then, that serotonin neurons may be computing the reliability of one of these processes in order to mediate their relative contributions. Such a mechanism would potentially result in an enhancement of model-based, faster learning around the time of probability transitions.

A number of studies have also examined the role of serotonin neuron activity in patience and persistence for rewards (Miyazaki et al., 2011, 2014; Fonseca et al., 2015; Lottem et al., 2018). These studies demonstrated that activating serotonin neurons increased waiting times for or active seeking of reward. In all cases, animals can be thought of as learning

from lack of rewards at each point in time. Under the proposed meta-learning framework, increasing expected uncertainty would slow this learning, resulting in prolonging waiting times or enhancing persistence.

What are the postsynaptic consequences of slow changes in serotonin release? Target regions involved in learning and decision making, like the prefrontal cortex, ventral tegmental area, and striatum, express a diverse range of serotonin receptors capable of converting a global signal into local changes in circuit dynamics. The activity in these regions also correlates with latent decision variables that update with each experience (Schultz et al., 1997; Samejima et al., 2005; Lau and Glimcher, 2008; Massi et al., 2018; Wang et al., 2018; Bari et al., 2019), providing a potential substrate through which serotonin could modulate learning. For example, the gain of RPE signals produced by dopamine neurons in the ventral tegmental area is modulated by the variance of reward value (Fiorillo et al., 2003; Tobler et al., 2005).

What is the presynaptic origin of uncertainty computation in serotonin neurons? Synaptic inputs from the prefrontal cortex (Geddes et al., 2016) may provide information about decision variables used in this task (Bari et al., 2019). Local circuit mechanisms in the dorsal raphe (Geddes et al., 2016; Zhou et al., 2017) and long-lasting conductances in serotonin neurons (Haj-Dahmane et al., 1991; Penington et al., 1993; Brown et al., 2002; Andrade et al., 2015; Gantz et al., 2020) likely contribute to the persistence of these representations.

For some time serotonin has been associated with behavioral flexibility without a mechanistic understanding of this relationship. Behaving flexibly requires modulating learning rates. In addition to explaining this aspect of behavior, the meta-learning RL model proposed here provides a quantitative link between serotonin neuron activity and flexible behavior. The model suggests that serotonin neurons track uncertainty to modulate learning rate.

3.4 Methods

Animals and surgery. We used 16 male and female mice, backcrossed with C57BL/6J and heterozygous for Cre recombinase under the control of the serotonin transporter gene (Slc6a4$^{tm1(cre)Xz}$, The Jackson Laboratory, 014554; Zhuang et al., 2005). 4 mice were used for electrophysiological recordings in the dynamic foraging task (4 male), 5 mice were used for electrophysiological recordings in the dynamic Pavlovian task (1 female, 4 male), and 7 mice (3 female, 4 male) were used for the chemogenetic experiments. Surgery was performed on mice between the ages of 4–8 weeks, under isoflurane anesthesia (1.0–1.5% in O_2) and in aseptic condtions. During all surgeries, custom-made titanium headplates were surgically attached to the skull using dental adhesive (C&B-Metabond, Parkell). After the surgeries, analgesia (ketoprofen, 5 mg kg^{-1} and buprenorphine, 0.05–0.1 mg kg^{-1}) was administered to minimize pain and aid recovery.

For electrophysiological experiments, we implanted a custom microdrive targeting dorsal raphe using a 16° posterior angle, entering through a craniotomy at 5.55 mm posterior to bregma and aligned to the midline.

For all experiments, mice were given at least one week to recover prior to water restriction. During water restriction, mice had free access to food and were monitored daily in order to maintain 80% of their baseline body weight. All mice were housed in reverse light cycle (12h dark/12h light, dark from 08:00–20:00) and all experiments were conducted during the dark cycle between 10:00 and 18:00. All surgical and experimental procedures were in accordance with the *National Institutes of Health Guide for the Care and Use of Laboratory Animals* and approved by the Johns Hopkins University Animal Care and Use Committee.

Behavioral task. Before training on the tasks, water-restricted mice were habituated to head fixation for 1–3 d with free access to water from the provided spouts (two 21 ga stainless steel tubes separated by 4 mm) placed in front of the 38.1 mm acrylic tube in which the mice rested. The spouts were mounted on a micromanipulator (DT12XYZ, Thorlabs) with a

custom digital rotary encoder system to reliably determine the position of the lick spouts in XYZ space with 5–10μm resolution (Bari et al., 2019). Each spout was attached to a solenoid (ROB-11015, Sparkfun) to enable retraction (see Behavioral tasks: dynamic foraging). The odors used for the cues (p-cymene and (−)-carvone) were dissolved in mineral oil at 1:10 dilution (30μl) and absorbed in filter paper housed in syringe adapters (Whatman, 2.7μm pore size). The adapters were connected to a custom-made olfactometer (Cohen et al., 2012) that diluted odorized air with filtered air by 1:10 to produce a 1.0 L min^{-1} flow rate. The same flow rate was maintained outside of the cue period so that flow rate was constant throughout the task.

Licks were detected by charging a capacitor (MPR121QR2, Freescale) or using a custom circuit (Janelia Research Campus 2019-053). Task events were controlled and recorded using custom code (Arduino) written for a microcontroller (ATmega16U2 or ATmega328). Water rewards were 2–4 μl, adjusted for each mouse to maximize the number of trials completed per session and to keep sessions around 60 minutes. Solenoids (LHDA1233115H, The Lee Co) were calibrated to release the desired volume of water and were mounted on the outside of the dark, sound-attenuated chamber used for behavior tasks. White noise (2–60 kHz, Sweetwater Lynx L22 sound card, Rotel RB-930AX two-channel power amplifier, and Pettersson L60 Ultrasound Speaker), was played inside the chamber to block any ambient noise.

Behavioral tasks: dynamic foraging. During the 1–3 days of habituation, mice were trained to lick both spouts to receive water. Water delivery was contingent upon a lick to the correct spout at any time. Reward probabilities were chosen from the set $\{0, 1\}$ and reversed every 20 trials.

In the second stage of training (5–12 d), the trial structure with odor presentation was introduced. Each trial began with the 0.5 s delivery of either an odor "go cue" ($P = 0.95$) or an odor "no-go cue" ($P = 0.05$). Following the go cue, mice could lick either the left or the right spout. If a lick was made during a 1.5 s response window, reward was delivered probabilistically from the chosen spout. The unchosen spout was retracted at the time of the

tongue contacting the other spout so that mice would not try to sample both spouts within a trial. The unchosen spout was replaced 2.5 s after cue onset. Following a no-go cue, any lick responses were neither rewarded nor punished. Reward probabilities during this stage were chosen from the set $\{0, 1\}$ and reversed every 20–35 trials. During this period of training only, water was occasionally manually delivered to encourage learning of the response window and appropriate switching behavior. Reward probabilities were then changed to $\{0.1, 0.9\}$ for 1–2 days of training prior to introducing the final stage of the task. Rewards were never "baited," as in previous versions of the task (Sugrue et al., 2004; Lau and Glimcher, 2005; Tsutsui et al., 2016; Bari et al., 2019). We did not penalize switching with a "changeover delay." If a directional lick bias was observed in one session, the lick spouts were moved horizontally 50–300μm in the opposite direction prior to the following session.

After the 1.5 s response window, inter-trial intervals were generated as draws from an exponential distribution with a rate parameter of 0.3 and a maximum of 30 s. This distribution results in a flat hazard rate for inter-trial intervals such that the probability of the next trial did not increase over the duration of the inter-trial interval (Luce, 1986). Inter-trial intervals (go-cue on to go-cue on) were 7.45 s on average (range 2.5–32.5 s). As in previous studies, mice made a leftward or rightward choice in greater than 99% of trials (Bari et al., 2019). Mice completed 280 ± 66.6 trials per session (range 79–655 trials).

In the final stage of the task, the reward probabilities assigned to each lick spout were drawn pseudorandomly from the set $\{0.1, 0.5, 0.9\}$ in all the mice from the behavior experiments ($n = 48$), all the mice from the DREADDs experiments ($n = 7$), and half of the mice from the electrophysiology experiments ($n = 2$). The other half of mice from the electrophysiology experiments ($n = 2$) were run on a version of the task with probabilities drawn from the set $\{0.1, 0.4, 0.7\}$. The probabilities were assigned to each spout individually with block lengths drawn from a uniform distribution of 20–35 trials. To stagger the blocks of probability assignment for each spout, the block length for one spout in the first block of each session was drawn from a uniform distribution of 6–21 trials. For each spout, probability assignments

could not be repeated across consecutive blocks. To maintain task engagement, reward probabilities of 0.1 could not be simultaneously assigned to both spouts. If one spout was assigned a reward probability greater than or equal to the reward probability of the other spout for 3 consecutive blocks, the probability of that spout was set to 0.1 to encourage switching behavior and limit the creation of a direction bias. If a mouse perseverated on a spout with reward probability of 0.1 for 4 consecutive trials, 4 trials were added to the length of both blocks. This procedure was implemented to keep mice from choosing one spout until the reward probability became high again.

For the DREADDs experiments, the probability of the task generating the special case probability transitions (Figure 3-5c-j) was enhanced when a new probability was being selected for one of the spouts at the end of a block. At this point, if one of the current probabilities was equal to 0.1 then we forced the special case transitions with $P = 1/3$. Medium ($P = 0.5$) and low ($P = 0.1$) were switched to low and high ($P = 0.9$), or high and low were switched to low and high simultaneously. Forced transitions were not allowed to occur in consecutive probability changes. This design increased the frequency of these transitions by $\sim$ 3x without drastically altering task structure or reward statistics.

To minimize spontaneous licking, we enforced a 1 s no-lick window prior to odor delivery. Licks within this window were punished with a new randomly-generated inter-trial interval, followed by a 2.5 s no-lick window. Implementing this window significantly reduced spontaneous licking throughout the entirety of behavioral experiments.

Behavioral tasks: dynamic Pavlovian. On each trial either an odor "CS+" ($P = 0.95$) or an odor "CS−" ($P = 0.05$) was delivered for 1 s followed by a delay of 1 s. CS+ predicted probabilistic reward delivery, whereas CS− predicted nothing. Mice were allowed 3 s to consume the water, after which any remaining reward was removed by a vacuum. Each trial was followed by an inter-trial interval, drawn from the same distribution as in the dynamic foraging task. The time between trials (CS on to CS on) was 9.34 s on average (range 6–36 s).

The reward probability assigned to CS+ was drawn pseudorandomly from the set $\{0.2, 0.5, 0.8\}$ or, in separate sessions, alternated between the probabilities in the set $\{0.2, 0.8\}$. The probability changed every 20–70 trials (uniform distribution). The CS+ probability of the first block of every session was 0.8.

Electrophysiology. We recorded extracellular signals from neurons at 32 or 30 kHz using a Digital Lynx 4SX (Neuralynx Inc,) or Intan Technologies RHD2000 system (with RHD2132 headstage), respectively. The recording systems were connected to 8–16 implanted tetrodes (32–64 channels, nichrome wire, PX000004, Sandvik) fed through 39 ga polyimide guide tubes that could be advanced with the turn of a screw on a custom, 3D-printed microdrive. The impedances of each wire in the tetrodes were reduced to 200–300 kΩ by gold plating. The tetrodes were wrapped around a 200 μm optic fiber used for optogenetic identification. After each recording session, the tetrode-optic-fiber bundle was driven down 75μm. The median signal was subtracted from raw recording traces across channels and bandpass-filtered between 0.3–6 kHz using custom MATLAB software. To detect peaks, the bandpass-filterd signal, x, was thresholded at $4\sigma_n$ where σ_n=median($\frac{|x|}{0.6745}$) (Quiroga et al., 2004). Detected peaks were sorted into individual unit clusters offline (Spikesort 3D, Neuralynx Inc.) using waveform energy, peak waveform amplitude, minimum waveform trough, and waveform principal component analysis. We used two metrics of isolation quality as inclusion criteria: L-ratio (< 0.05) (Schmitzer-Torbert et al., 2005) and fraction of interspike interval violations ($< 0.1\%$ interspike intervals < 2 ms).

Individual neurons were determined to be optogenetically-identified if they responded to brief pulses (10 ms) of laser stimulation (473 nm wavelength) with short latency, small latency variability, and high probability of response across trains of stimulation (10 trains of 10 pulses delivered at 10 Hz). We used an unsupervised k-means clustering algorithm to cluster all neurons based on these features. The elbow method and Calinkski-Harabasz criterion were used to determine that the optimal number of clusters was 4. Members of the cluster (66 neurons) with the highest mean probability of response, shortest mean latency,

and smallest mean latency standard deviation were considered as identified. The responses of individual neurons were manually inspected to ensure light responsivity. In addition to the presence of identified serotonin neurons, targeting of dorsal raphe was confirmed by performing electrolytic lesions of the tissue (20 s of 20 μA direct current across two wires of the same tetrode) and examining the tissue after perfusion.

Viral injections. To express channelrhodopsin-2 (ChR2), hM4di, or mCherry in dorsal raphe serotonin neurons, we pressure-injected 810 nl of rAAV5-EF1a-DIO-hChR2(H134R)-EYFP (3×10^{13} GC ml^{-1}),
pAAV-hSyn-DIO-hM4D(Gi)-mCherry (1.2×10^{13} GC ml^{-1}), or pAAV-hSyn-DIO-mCherry (1.0×10^{13} GC ml^{-1}) into the dorsal raphe of *Sert*-Cre mice at a rate of 1 nl/s (MMO-220A, Narishige). pAAV-hSyn-DIO-hM4D(Gi)-mCherry was a gift from Bryan Roth (Addgene viral prep 44362-AAV5). We made three injections of 270 nl at the following coordinates: $\{4.63, 4.57, 4.50\}$ mm posterior of bregma, $\{0.00, 0.00, 0.00\}$ mm lateral from the midline, and $\{2.80, 3.00, 3.25\}$ mm ventral to the brain surface. The pipette was inserted through a craniotomy at -5.55 mm posterior to bregma and aligned to midline, using a 16° posterior angle. Before the first injection, the pipette was left at the most ventral coordinate for 10 minutes. After each injection, the pipette was withdrawn 50μm and left in place for 5 min. The craniotomy after a hM4Di or mCherry injection was covered with silicone elastomer (Kwik-Cast, WPI) and dental cement. For electrophysiology experiments with rAAV5-EF1a-DIO-hChR2(H134R)-EYFP injections, the microdrive was implanted through the same craniotomy.

Inactivation of serotonin neurons. 4 mice were injected with pAAV-hSyn-DIO-hM4D(Gi)-mCherry and 4 mice were injected with pAAV-hSyn-DIO-mCherry as a control. One of the hM4D mice failed to perform the task and so was excluded. After training mice, we injected either 3.0 mg kg^{-1} agonist 21 (Tocris) dissolved in 0.5% DMSO/saline or an equivalent volume of vehicle (0.5% DMSO/saline alone) I.P. on alternating days (5 sessions per injection type per mouse).

Data analysis. All analyses were performed with MATLAB (Mathworks). All data are presented as mean ± S.D. unless reported otherwise. All statistical tests were two-sided. In Figure 2-3d, the probability that the time constants from the actual behavior belonged to the distribution of simulated behavior time constants was calculated by finding the Mahalanobis distance of the former from the latter, calculating the cumulative density function of the chi-square distribution at that distance, and subtracting it from 1. For all analyses, no-go (dynamic foraging) and CS− (dynamic Pavlovian) cues were ignored and treated as part of the inter-trial interval.

Data analysis: generative model of behavior with meta-learning. We observed mouse behavior that the static learning model failed to capture and that suggested that learning rate was not constant over time. Thus, we added a component to the model that modulates RPE magnitude and $\alpha_{(-)}$ ("meta-learning"). Because learning should be slow in stable but variable environments, expected uncertainty scaled RPEs, such that learning is decreased when expected uncertainty is high. If the left spout was chosen, the values of actions were updated according to

$$Q_l(t+1) = \begin{cases} Q_l(t) + \alpha_{(+)}\delta(t)(1-\epsilon(t)), & \text{if } \delta(t) > 0 \\ Q_l(t) + \alpha_{(-)}(t)\delta(t)(1-\epsilon(t)), & \text{if } \delta(t) < 0 \end{cases}$$
$$Q_r(t+1) = \zeta Q_r(t),$$

where ϵ is an evolving estimate of expected uncertainty calculated from the history of unsigned RPEs:

$$\upsilon(t) = |\delta(t)| - \epsilon(t),$$
$$\epsilon(t+1) = \epsilon(t) + \alpha_\upsilon \upsilon(t).$$

The rate of RPE magnitude integration is controlled by α_υ. Deviations from the expected uncertainty are captured by unexpected uncertainty, υ, and may indicate that a change has

occurred in the environment. Changes in the environment should drive learning to adapt behavior to new contingencies so $\alpha_{(-)}$ varies as a function of how surprising recent outcomes are:

$$\alpha_{(-)}(t) = \begin{cases} \alpha_{(-)}(t-1) & \text{if } \delta(t) > 0 \\ \psi(\upsilon(t) + \alpha_{(-)_0}) + (1-\psi)(\alpha_{(-)}(t-1)) & \text{if } \delta(t) < 0 \end{cases}$$

where $\alpha_{(-)_0}$ is the baseline learning rate from no reward and ψ controls how quickly unexpected uncertainty is integrated to update $\alpha_{(-)}$. As it is formulated, $\alpha_{(-)}$ increases after surprising no reward outcomes. This learning rate was not allowed to be less than 0, such that

$$\alpha_{(-)}(t) = 0, \text{if } \alpha_{(-)}(t) < 0$$

The Q-values are used to generate choice probabilities through a softmax decision function (Daw et al., 2006):

$$P(c(t) = r) = \frac{1}{1 + e^{-\beta(Q_r(t) - Q_l(t) + bias)}},$$
$$P(c(t) = l) = 1 - P(c(t) = r),$$

where β, the "inverse temperature" parameter, controls the steepness of the sigmoidal function. In other words, β controls the level of exploration versus exploitation with respect to the relative action values.

Data analysis: firing rate model. We developed a version of our meta-learning model to fit inter-trial firing rates to see if neural activity and choice behavior reported similar dynamics of expected uncertainty. The learning components of the models were identical, but the firing rate model fit z-scored firing rates as a function of expected uncertainty:

$$\mu(t) = slope \cdot \bar{v} + intercept,$$
$$FR(t) \sim \mathcal{N}(\mu(t), \sigma),$$

where *slope* and *intercept* scale expected uncertainty into the mean predicted firing rate, μ. Real firing rates, FR, are modeled as a draw from a Gaussian distribution with mean μ and some fixed amount of noise, σ.

Data analysis: Model fitting. We fit and assessed models using MATLAB (Mathworks) and the probabilistic programming language, Stan (https://mc-stan.org/) with the MATLAB interface, MatlabStan (https://mc-stan.org/users/interfaces/matlab-stan). Stan was used to construct hierarchical models with mouse-level hyperparameters to govern session-level parameters. For each session, each parameter in the model (for example, α_ϵ for the meta-learning model) was modeled as a draw from a mouse-level distribution with mean μ and variance σ. Models were fit using noninformative (uniform distribution) priors for session-level parameters ($[0, 1]$ for all parameters except β which was $[0, 10]$) and weakly informative ($\mu \sim \mathcal{N}(0, 1)$, $\sigma \sim$ half-Cauchy$(0, 3)$) priors for mouse-level hyperparameters. For some mice with fewer sessions, more informative mouse-level hyperparmeters were used to achieve model convergence under the assumption that individual mice behave similarly across days. This hierarchical construction mitigated the typical variability of point estimates for session-level parameters that results from other methods of estimation. Stan uses full Bayesian statistical inference to generate posterior distributions of parameter estimates using Hamiltonian Markov chain Monte Carlo sampling (Carpenter et al., 2017). The parameters for updating expected uncertainty, α_υ, and for updating the negative RPE learning rate, ψ, were ordered such that $\psi > \alpha_\upsilon$. The ordering operated under the assumption that learning rate should be integrated more quickly to detect change. The ordering also helped models to converge more quickly.

Data analysis: extracting model parameters and variables, behavior simulation.

For extracting model variables (like expected uncertainty), we took at least 4,000 draws from the Hamiltonian Markov Chain Monte Carlo samples of session-level parameters, ran the model agent through the task with the actual choices and outcomes, and averaged each model variable across runs. For comparisons of individual parameters across behavioral and neural models, we obtained maximum *a posteriori* parameter values by approximating the mode of the distribution: binning the values in 50 bins and taking the median value of the most populated bin. For simulations of behavior, we took at least 4,000 draws from the Hamiltonian Markov Chain Monte Carlo samples of mouse-level parameters and simulated behavior and outcomes in a number of random sessions per sample. For the transition analysis, that number was proportional to the number of rare transitions that each animal contributed to the actual data. For other analyses that number was fixed.

Data analysis: linear regression models of neural activity. For comparisons of firing rates to the behavioral-model-generated uncertainty terms we regressed z-scored firing rates on z-scored uncertainty using the MATLAB function "fitlm". For some neurons and sessions, firing rates and model variables demonstrated monotonic changes across the session. To control for the effect of these dynamics in comparisons of inter-trial interval firing rates to model variables, we regressed out the monotonic effects for each term separately, then regressed the firing rate residuals on the expected uncertainty residuals. Here, we found similar rates of correlation across the population of neurons. We also looked for relationships between the neural activity and other model variables that evolved as a function of action and outcome history. For the analysis of the dynamic foraging task data, we added total value ($Q_r + Q_l$), relative value ($Q_r - Q_l$), value confidence ($|Q_r - Q_l|$), RPE, and reward history as regressors in the same model (Figure A3-2e). Value confidence captures how much better the better option is on each trial. Reward history is an arbitrarily smoothed history of all rewards, generated by convolving rewards with a recency-weighted kernel. The kernel was derived from an exponential fit to the coefficients from the regression of choices on outcomes. For the dynamic Pavlovian task data, we added RPE and reward history as regressors (Figure

3-4e).

Data analysis: linear mixed effect models. To analyze the changes in transition behavior we constructed a linear mixed effects model that predicted choice averages after transition points as a function of trial since transition, transition type, and the interaction between the two. The model is described by the following Wilkinson notation:

$$choice\,averages \sim trial\,from\,transition * transition\,type.$$

For assessing the affect of chemogenetic manipulation, we constructed a linear mixed effects model that predicted choice averages after transition points as a function of trial since transition, transition type, and the interaction between the two. We added drug condition (vehicle or agonist 21) as a fixed effect as well as the interaction between transition type and drug condition:

$$choice\,averages \sim trial\,from\,transition * transition\,type + transition\,type * drug\,condition.$$

In the case of simulated data, these fixed effects were grouped by mouse, treated as a random effect that affects both slope and intercept, given by:

$$choice\,averages \sim trial\,from\,transition * transition\,type + transition\,type * drug\,condition + (trial\,from\,transition * transition\,type \mid mouse) + (transition\,type * drug\,condition \mid mouse).$$

In all models, we z-scored all choice probabilties to center the data.

Data analysis: autocorrelation controls. To control for potential statistical confounds in correlating two variables with similar autocorrelation functions — in particular, firing rates of serotonin neurons and dynamics of expected uncertainty — we simulated each variable and compared it to the real data. We simulated 1,000 expected uncertainty variables by using maximum *a posteriori* parameter estimates to simulate a random sequence of choices and outcomes of the same length as the real session. For each simulation we extracted model variables using the sampling and averaging method described above. Linear regressions of real

firing rates on each simulated variable were performed. If the t-statistics from the regression of real firing rate on real model variable fell beyond the 95% boundary of the distribution of t-statistics from the comparisons with simulated variables, then the relationship was deemed significant. We view this control analysis as an estimate of a lower bound on the true rate of correlated variables; for example, in a recent paper, only approximately one-third of true correlations were recoverable with this simulation (Elber-Dorozko and Loewenstein, 2018).

Conversely, we simulated neural data with autocorrelation functions matched to those of the actual neuron. For each neuron, we computed the autocorrelation function for lags of 10 trials and calculated the sum. The autocorrelation function sum was mapped onto the scale of a half-Gaussian smoothing kernel (width of 10 trials) using a log transformation. Neurons were then simulated as a random walk such that the firing rate at a given trial was the sum of the previous 10 trials weighted by the smoothing kernel plus some normally distributed noise ($\mathcal{N}(0,1)$). We found that the autocorrelation functions and the distributions of simulated firing rates were similar to those of the real neurons. For each real neuron, we performed 1,000 simulations and compared them to the real expected uncertainty in the same way as described above.

Histology. After experiments were completed, mice were euthanized with an overdose of isoflurane, exsanguinated with saline, and perfused with 4% paraformaldehyde. The brains were cut in 100-μm-thick coronal sections and mounted on glass slides. We validated expression of rAAV5-EF1a-DIO-hChR2(H134R)-EYFP, pAAV-hSyn-DIO-hM4D(Gi)-mCherry, or pAAV-hSyn-DIO-mCherry with epifluorescence images of dorsal raphe (Zeiss Axio Zoom.V16). In electrophysiological experiments, we confirmed targeting of the optic-fiber-tetrode bundle to the dorsal raphe by location of the electrolytic lesion.

Chapter 4

Serotonin neurons may modulate learning rate in medial prefrontal cortex

Abstract

Dorsal raphe serotonin neurons modulate how quickly mice learn from the consequences of their actions. In doing so, these neurons may compute estimates of uncertainty. However, this relationship is not present in the activity of the entire population, suggesting that only certain subtypes or pathways are responsible for disseminating this information. In the framework of some normative and meta-learning reinforcement learning models, uncertainty modulates how quickly the expected values of actions are updated. Previous work demonstrated that the activity of medial prefrontal cortex corresponds to the value of actions in mice behaving in a dynamic foraging task. Using a similar foraging task, we manipulated dorsal raphe serotonin inputs to this region to test the hypothesis that this pathway modulates how quickly those value representations are updated. Preliminary results suggest that optogenetic activation of serotonin axon terminals may change dynamics of value-related firing rates in medial prefrontal cortex. Effects on behavior are consistent with serotonin neuron modulation of learning rates through enhancement of expected uncertainty.

4.1 Introduction

In order to behave in a way that maximizes value, the brain is constantly learning about correlational relationships between actions and the outcomes they tend to produce. This task is a difficult one, especially when those relationships are noisy and subject to change. Normative models propose that behavior can be optimized by modulating learning rate as a function of uncertainty (Dayan et al., 2000; Yu and Dayan, 2005; Soltani and Izquierdo, 2019; Preuschoff et al., 2006; Preuschoff and Bossaerts, 2007; Diederen and Schultz, 2015). Expected uncertainty about these relationships should mitigate learning so that behavior is not driven suboptimally by noisy outcomes. Unexpected uncertainty should enhance learning in order to respond appropriately to a perceived change in the action-outcome contingency. Human and mouse behavior can be characterized by models that adapt learning rates in this way (Nassar et al., 2012; McGuire et al., 2014; Kao et al., 2020). In humans, the rate of learning about the errors of predictive inference is modulated by the variance of outcomes (Nassar et al., 2012; McGuire et al., 2014; Kao et al., 2020). When subjects try to guess a number produced by a distribution, how quickly they learn from the discrepancy between their guess and the number is a function of the variance of that distribution(Nassar et al., 2012). These learning rates are also enhanced when a change is detected in the mean of that distribution.

We previously showed that mice also exhibit variable learning rates during dynamic foraging (Chapter 2). This behavior can be characterized by uncertainty-driven learning in a similar manner. In our meta-learning model the brain learns an estimate of expected uncertainty by calculating a moving average of unsigned prediction errors. In other words, this is an estimate of how far off, on average, were expected values from the actual outcomes. This estimate, we found, generalizes across actions and so may represent uncertainty about a behavioral policy. Deviations from this estimate (unexpected uncertainty) are integrated to drive learning from less-than-expected outcomes. We should note that this negative-

error-specific modulation of learning may be a consequence of task design and learning from greater-than-expected outcomes may be modulated as well.

In this work, the activity of roughly half of recorded dorsal raphe serotonin neurons correlated with estimates of uncertainty from a meta-learning reinforcement learning model (Chapter 3). Consistent with previous observations of heterogeneous responses to task and behavioral events (Nakamura et al., 2008; Ranade and Mainen, 2009; Li et al., 2013; Hayashi et al., 2015; Cohen et al., 2015), both positive and negative correlations were demonstrated. Inhibition of serotonin neurons also affected learning in accordance with a proposed role in encoding uncertainty. Given the heterogeneity, it is possible that uncertainty is only being signaled to certain downstream targets to modulate learning.

Dorsal raphe serotonin neurons innervate the majority of the forebrain, including distributed regions associated with learning and decision making (Jacobs and Azmitia, 1992). One region that may implement changes in learning as a function of serotonin neuron input is the medial prefrontal cortex (mPFC). Serotonin neurons from dorsal raphe make dense axonal arborizations in mPFC (Jacobs and Azmitia, 1992; Linley et al., 2013; Prouty et al., 2017; Ren et al., 2018) and serotonin modulates local cell and circuit function (Avesar and Gulledge, 2012; Stephens et al., 2014; Kjaerby et al., 2016; Athilingam et al., 2017). Neural activity in this region of the brain also relates to the value of actions and uncertainty-driven decision making. In the predictive inference task described above, converging correlations between fMRI BOLD signals and uncertainty as well as change point probability were observed across regions in prefrontal cortex, including dorsomedial prefrontal cortex (McGuire et al., 2014). In a similar task, activity in regions of the prefrontal cortex, which also included dorsomedial prefrontal cortex, predicted choice as a function of uncertainty-modulated error magnitude (Kao et al., 2020). In a mouse dynamic foraging task, the activity of single mPFC neurons correlated with decision variables from a reinforcement learning model (Bari et al., 2019). Specifically, neural activity reflected the relative value of two available actions as well as the sum of the action values over long timescales. It was also shown that this activity was

necessary for flexible decision making behavior. Incoming signals related to uncertainty from serotonin neurons then, may modulate these representations.

Here, we record action potentials from single neurons in mPFC of mice behaving in a dynamic foraging task that produces adaptive learning and decision making behavior. Confirming our previous results, this behavior is well-characterized by a meta-learning reinforcement learning model that uses uncertainty to guide adaptive learning. Using this model and a version with static learning rates, we also confirm the presence of decision variable representations in mPFC (Bari et al., 2019). Optogenetic activation of serotonin axons in the recorded region were used to test the hypothesis that serotonin neuron activity modulates how quickly value representations are updated in order to drive adaptive behavior. Preliminary results show changes in behavior and firing rates that are consistent with an enhancement of expected uncertainty.

4.2 Results

mPFC single neuron activity reflects action values

Mice (n = 5; 3 female and 2 male) implanted with tetrodes unilaterally and optic fibers bilaterally in mPFC were trained on the same dynamic foraging task described in Chapter 2. In accordance with previous work, we fit a reinforcement learning model with static learning rates to examine representations of decision variables in single neurons (Bari et al., 2019). The first variable is the difference between the expected values of the two alternative choices ($Q_r - Q_l$). In the formulation of the model, this relative value determines the probability that an animal will make a certain choice. The second variable is the sum of the expected values ($Q_r + Q_l$). This total value term may inform behavioral vigor, as it is inversely correlated with response latency (Bari et al., 2019). We used generalized linear models (Poisson regressions) to model spikes during the last 1 s of the intertrial interval (ITI) as a function of relative and total value. This period of time was selected because no licks can occur during this

window (otherwise a delay is enforced) and the activity is less likely to be contaminated by the phasic activity seen within the trial. Although there are differences in the reward statistics imposed by task structure across studies, we found similar rates of decision variable correlations as found previously (Bari et al., 2019). Many neurons had significant regression coefficients for one or both of relative value and total value (319 of 363, 88%). Some neurons also had significant regression coefficients exclusively ("pure" neurons) for relative value (55 of 363, 15%) or total value (80 of 363, 22%). For pure relative value neurons we found both right-preferring ($Q_r - Q_l$) and left-preferring ($Q_l - Q_r$) neurons. For pure total value neurons, we found both positive and negative correlations as well.

Consistent with the observations detailed in Chapter 2, we found evidence that mice use variable learning rates. Specifically, choice average curves around certain reward probability changes indicated a difference in learning rate as a function of recent outcome statistics (Figure 4-1a). This behavior was captured by our meta-learning model. Using decision variables from the meta-learning model, we used the same regression model approach to characterize the relationship between ITI firing rates and decision variables. Results were similar to the regression models using decision variables from the static learning model. Again, many neurons had significant relationships with one or both of the variables (322 of 363, 89%). Pure neurons also had significant relationships with relative value (60 of 363, 17%; Figure 4-1b-d) or total value (80 of 363, 22%; Figure 4-1e-g).

Activation of dorsal raphe serotonin neuron axons may modulate mPFC activity

Having confirmed the widespread representation of decision variables in mPFC, we sought to address the hypothesis that dorsal raphe serotonin neuron inputs to the region modulate the rate at which those variables are updated. Channelrhodopsin-2 was expressed in dorsal raphe serotonin neurons under the control of the serotonin transporter promoter in *Slc6a4*-Cre (also known as *Sert*-Cre) mice. Optogenetic activation of serotonin neuron axons in mPFC has been

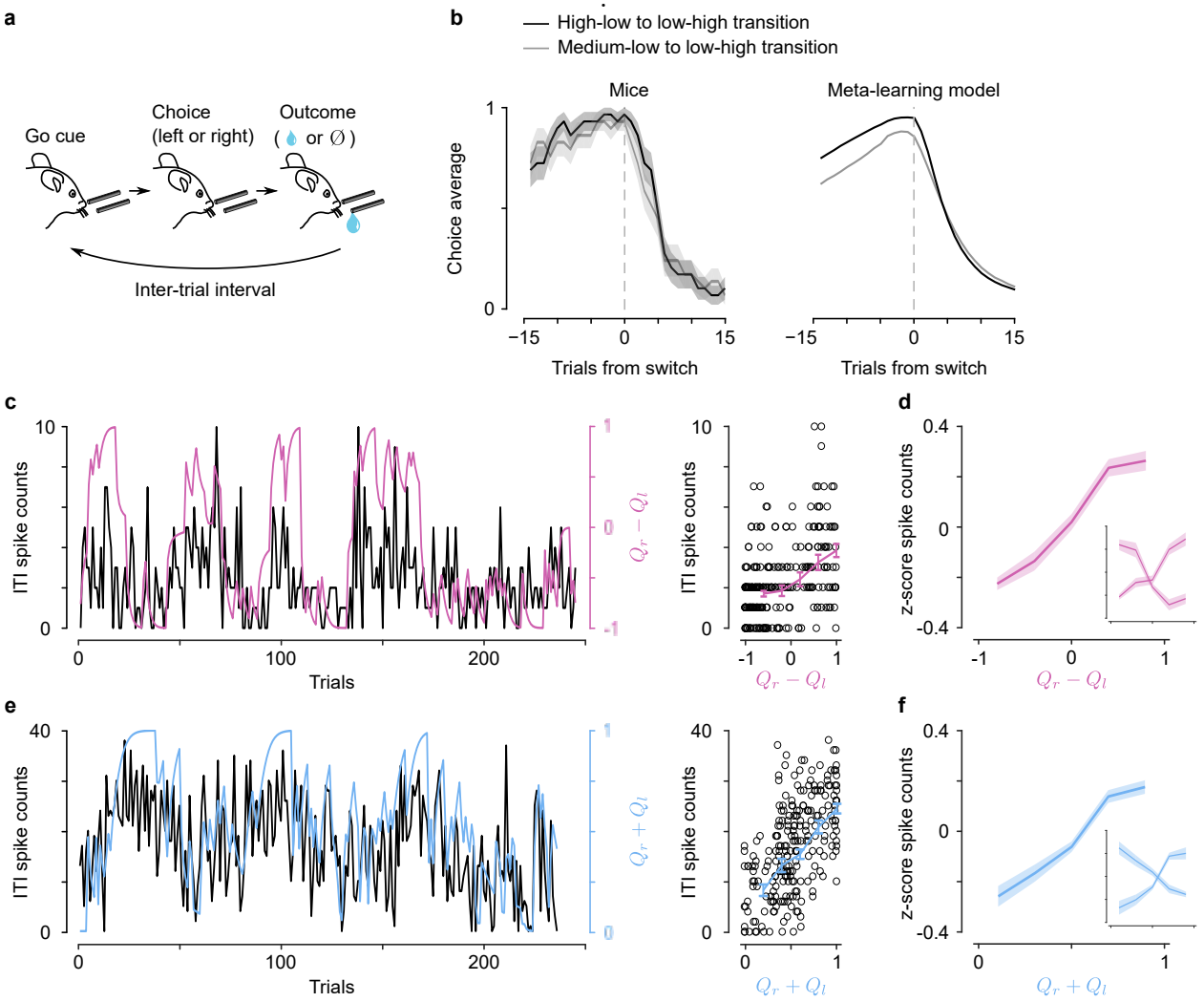

Figure 4-1: **mPFC neurons track decision variables during meta-learning.** (a) Dynamic foraging task in which mice chose freely between a leftward and rightward lick, followed by a reward with a probability that varied over time. (b) Left: actual mouse behavior at transitions in which reward probabilities change simultaneously ($n = 29$ high-low to low-high, $n = 29$ medium-low to low-high). Lines are mean choice probability relative to the spout that initially has the higher probability, shading is Bernoulli SEM. Right: simulated behavior at transitions using meta-learning model parameters fit to actual behavior. (c) Example pure relative value neuron. Left: Spike counts during the ITI (black) for each trial compared to relative value estimates (pink) from the meta-learning model. Right: scatter plot of the same data (black circles), with average spike counts from bins of relative values (pink line). (d) Population z-scored firing rates varied with relative value. Inset shows population split by positive and negative correlations. Main plot combines across neurons by "sign-flipping" the left-preferring neurons. (e) Same as (c) but for a pure total value neuron. (f) Same as (d) but for total value neurons.

shown to elicit serotonin release *in vitro* (Sparks et al., 2017; Athilingam et al., 2017). Blue light was delivered at 50 Hz during the odor go-cue (500 ms). These stimulation parameters were chosen to artificially enhance the phasic representation of expected uncertainty seen in recordings from identified serotonin neurons during this epoch. On alternating days, the patch cords delivering the light rested near the implant as a sham control or were connected to the optic fibers bilaterally implanted in mPFC to activate serotonin neuron axons. Therefore, on activation days, the light was delivered in the region whose activity was sampled by the tetrodes and the contralateral mPFC. We advanced the tetrodes at the end of the session on alternating days, with the intent to sample approximately the same population of neurons for each condition.

We first looked for neural evidence that serotonin axon activation modulated firing rates. To rule out simple inhibition or excitation across the population, we examined basic firing rate properties for all neurons across conditions. We found no significant effects of stimulation on average firing rate ($T_{707} = 1.02$, $p = 0.31$), firing rate variance ($T_{707} = -0.179$, $p = 0.86$), nor on firing rates specifically during the go-cue ($T_{707} = 0.341$, $p = 0.73$), outcome ($T_{707} = 1.43$, $p = 0.15$), or last second of the ITI ($T_{707} = 0.858$, $p = 0.39$).

These basic analyses do not speak to potential effects on relationships between activity and task or behavioral variables since those relationships may be present only in specific neurons and with opposite changes in firing rates. To get some insight into how stimulation could be modulating responses to observable variables, we performed a series of linear regressions. For each neuron we compared firing rates to outcome, choice, choice-outcome interaction, and previous outcome in 500 ms sliding windows (moved in 100 ms increments) throughout the trial (Figure 4-2a). At the population level, we observed modest changes in correlations. Of note, there was a decrease in the fraction of neurons with significant regression coefficients for outcome around the time of the outcome. There was also a decrease in the fraction of neurons whose activity was related to previous outcome, throughout the trial period.

These modest changes may indirectly reflect more meaningful changes in how outcomes

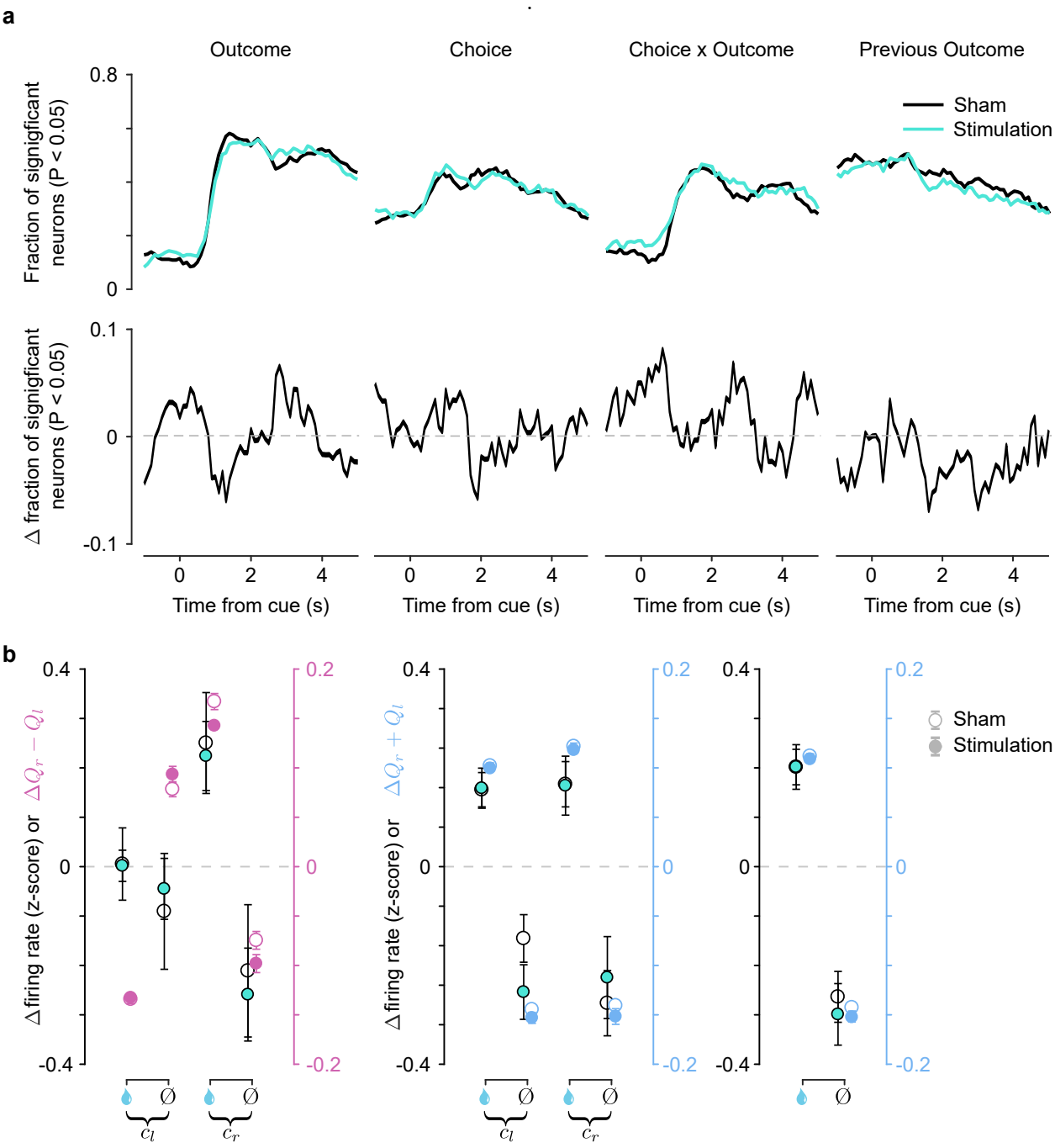

Figure 4-2: **Serotonin axon stimulation may affect task responsivity of mPFC neurons.** (a) Top row: Results from linear regression models predicting firing rates of individual neurons, at different timepoints, as a function of observable variables. Results are shown for sham and stimulation conditions. Bottom row: Small differences in population rates of observable variable encoding. (b) Comparisons of changes in firing rate (black circles) with changes in relative value (pink circles) or total value (blue circles) as a function of reward (water droplet) or no reward (∅) given a left choice (c_l) or right choice (c_r). Changes are compare across sham (empty circles) and stimulation conditions (filled circles). Sign-flipping was used to pool across neurons.

are used to update decision variables. In examining this possibility, we fit the meta-learning model to behavior in each condition separately and compared changes in decision variables to changes in firing rates (Figure 4-2b). Q-learning models make specific predictions about how decision variables will change as a function of choice and outcome (Bari et al., 2019). Total value increases as a function of rewards and decreases as a result of no rewards, regardless of choice. Relative value, on the other hand, increases as a consequence of rewards on the preferred side as well as no rewards on the non-preferred side. Relative value decreases with the opposite action and outcome pairings.

Analysis of changes in firing rate was restricted to pure neurons. For relative value neurons, we sign-flipped the left-preferring neurons so that they were fictively right-preferring in order to average across neurons. We used the analogous strategy for pooling pure total value neurons as well. Despite these strategies, the preliminary results are still underpowered. Part of this limitation is a result of not standardizing firing rates, since we were comparing across conditions. However, the effects on firing rate changes are almost all trending in the same direction as the effects on model variables.

Activation of dorsal raphe serotonin neuron axons in mPFC changes behavior, learning rates, and uncertainty

The previous analysis suggested an effect of stimulation on learning, but did not take into account outcome history in looking at changes in model variables. Further, the meta-learning model we have specified has variable learning rates, so we would expect updating to not be uniform within action-outcome pairings. To get clearer insight into potential changes in learning as a result of stimulation, we examined model parameters across mice as well as model variables throughout sessions. Mode estimates of the parameters that controlled meta-learning mostly increased, but the change was not significant across the group (Figure 4-3a). These parameters control the updating rate of expected uncertainty (α_v and the integration rate of unexpected uncertainty (psi) that determines learning rate from less-than-expected

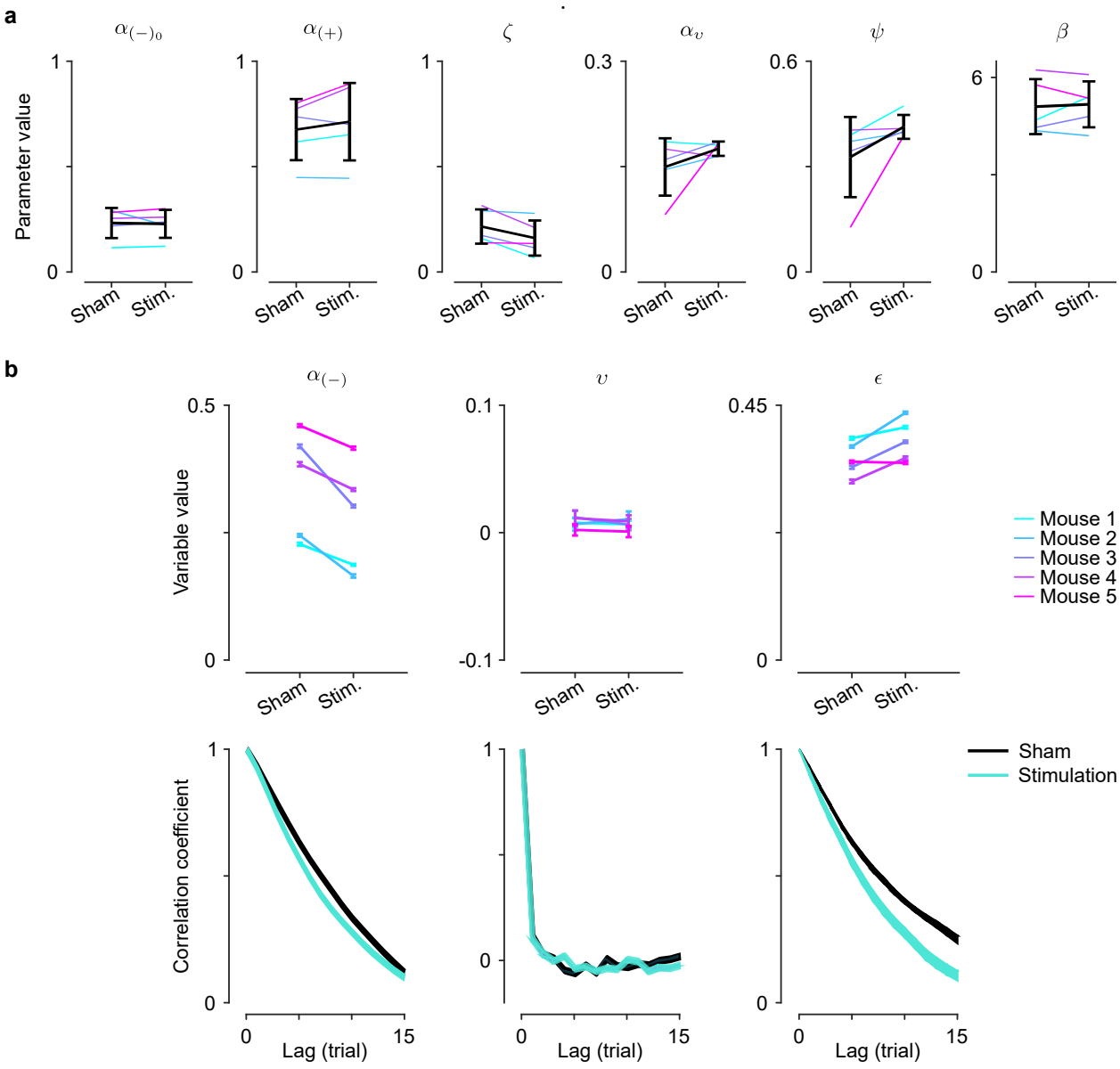

Figure 4-3: **Serotonin axon stimulation enhances expected uncertainty and attenuates learning.** (a) Maximum *a posteriori* (mode) estimates of parameters from the meta-learning model vary across manipulation conditions. Colored lines indicate individual mice and black lines represent mean ± SEM. From left to right: median negative RPE learning rate ($\alpha_{(-)_0}$), positive RPE learning rate ($\alpha_{(+)}$), forgetting rate (ζ), expected uncertainty update rate (α_υ), unexpected uncertainty integration rate (ψ), and the inverse temperature of the decision function (β). (b) Top row: changes in model variables across manipulation conditions. Bottom row: autocorrelation functions for model variables. From left to right: negative RPE learning rate ($\alpha_{(-)}$), unexpected uncertainty (υ), and expected uncertainty (ϵ).

outcomes. Changes in the forgetting rate parameter consistently decreased, indicating that the expected value of the unchosen spout was devalued more quickly.

Within animal there were small changes across multiple parameters. Because model parameters interact in driving model variables, we also looked at changes in the latter. Interestingly, we found that expected uncertainty increased significantly in the stimulation condition (Figure 4-3b). Additionally, there was smaller autocorrelation in the variable, suggesting that expected uncertainty was being updated more quickly. Learning rates for less-than-expected outcomes decreased significantly, but also became more variable as a consequence of faster updating.

We also looked at behavior during simultaneous probability switches to see how these changes in variables affected observable meta-learning. There was a substantial decrease in the speed of adaptation when medium and low probabilities changed to low and high (Figure 4-4a). This change was partially captured by the meta-learning model (Figure 4-4b). Model variable averages were plotted to get a better sense of how they evolve around transitions across probability and manipulation conditions (Figure 4-5). In the transition from high and low to low and high, stimulation enhanced learning from negative RPEs initially, but this effect was mitigated by a more rapid increase in expected uncertainty. In the other reward probability condition, however, the decrease in learning after the transition seen in the sham condition was attenuated in the stimulation condition. In other words, learning from negative RPEs was enhanced. Combined with higher expected uncertainty in this reward probability condition (which attenuates learning from both rewards and no rewards), this change in learning from no rewards may drive suboptimal switching away from the better side.

4.3 Discussion

Expected and unexpected uncertainty can be leveraged to modulate learning rates in order to produce flexible behavior in noisy and dynamic environments. We provide evidence that

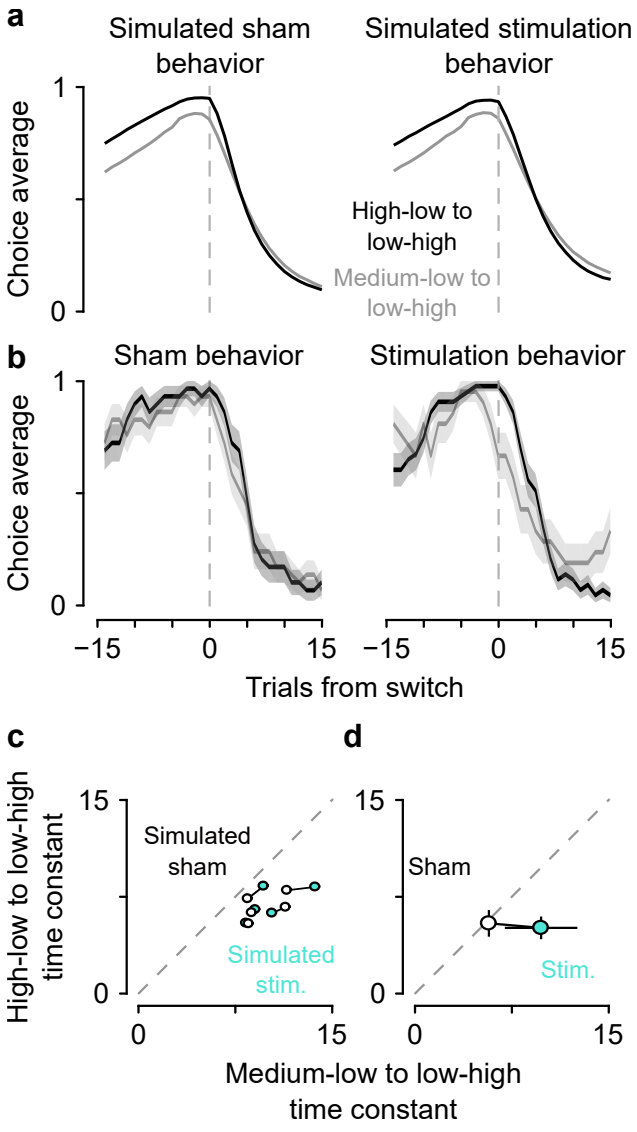

Figure 4-4: **Serotonin axon stimulation modulates observable meta-learning.** (a) Simulated behavior with meta-learning model, fit to sham behavior (left) and stimulated behavior (right). (b) Mouse behavior in the sham and stimulation conditions. Lines are mean choice probability and shading is Bernoulli S.E.M. (c) Exponential time constants for transitions from simulated behavior across conditions. (f) Time constants from mice (with 95% C.I.).

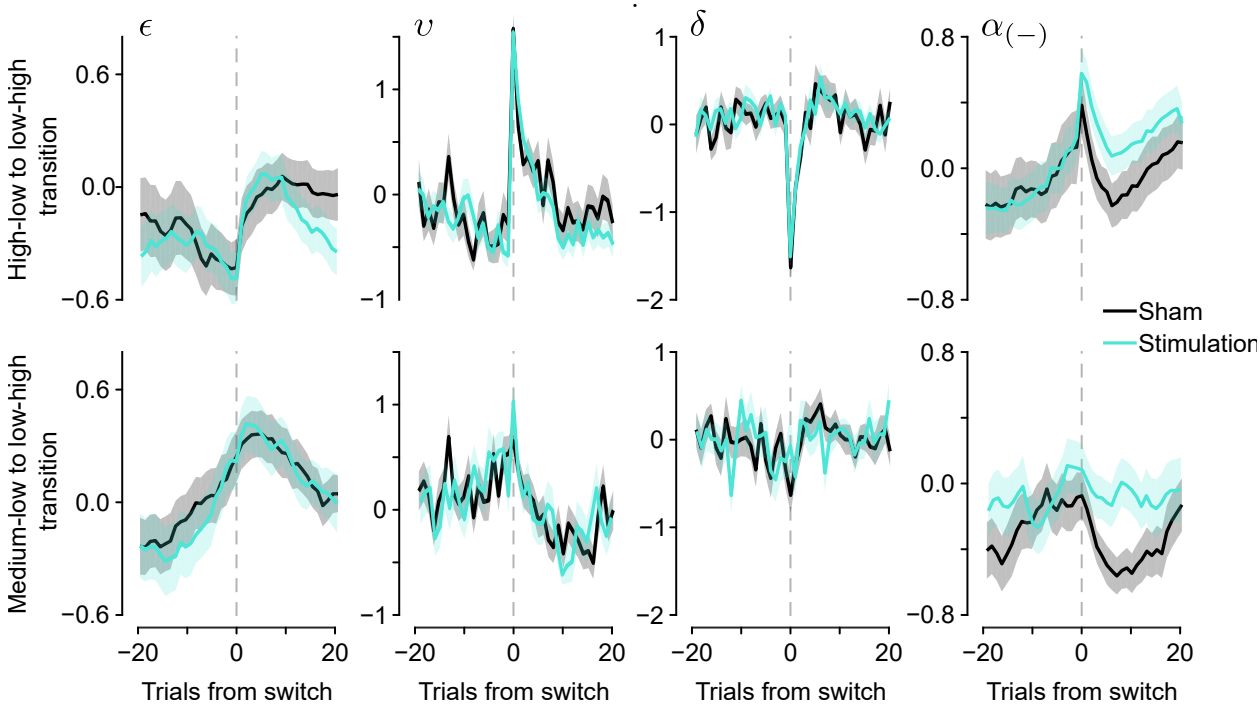

Figure 4-5: **Serotonin axon stimulation affects meta-learning model variables.** Trial-by-trial dynamics of z-scored values of expected uncertainty (ϵ), unexpected uncertainty (υ), reward prediction error (δ), and negative learning rate ($\alpha_{(-)}$) around transitions in reward probabilities (cf. Figure 4-4a) for sham and stimulation conditions. Mean ± S.E.M. z-scored values are plotted for each variable.

serotonin neurons may provide signals of uncertainty to mPFC in order to modify learning and decision making. We confirmed mPFC representations of relative and total value using decision variables from static learning and meta-learning reinforcement learning models. Estimates of these decision variables changed as a function of optogenetic activation of local serotonin neuron axons. Estimates of expected uncertainty were enhanced across sessions as a consequence of stimulation, consistent with the hypothesis that phasic serotonin neuron activity during the go-cue is related to the variable. Expected uncertainty was also more variable which may be due to faster update rates. General increases in expected uncertainty led to smaller learning rates from negative RPEs. These learning rates were also more variable, which may also be due to faster integration of unexpected uncertainty.

Variability in expected uncertainty and negative RPE learning rates might also be a result of the model trying to capture the artificial changes in these signals brought about by optogenetic activation. Subsequent analyses will model this possibility explicitly, fitting control and stimulation behavior simultaneously with an additional parameter that adds to expected uncertainty in the stimulation condition.

The mPFC may be the substrate of learning changes mediated by serotonin, but it may also provide uncertainty-related signals to serotonin neurons. Dorsal raphe neurons receive considerable input from mPFC neurons (Pollak Dorocic et al., 2014; Weissbourd et al., 2014; Challis et al., 2014; Ogawa et al., 2014). These inputs drive serotonin neurons directly through glutamatergic synapses, then inhibit them disynaptically via activation of local GABA neurons (Geddes et al., 2016; Zhou et al., 2017). An interesting possibility is that this circuit provides the mechanism by which expected uncertainty is phasically encoded during expectation and subtracts this expectation during the outcome period. This subtraction would allow for unexpected uncertainty to be computed at the time of the outcome. A similar encoding scheme has been observed in ventral tegmental area dopamine and GABA neurons Cohen et al. (2012); Eshel et al. (2015). In recordings of serotonin neurons, phasic activation during the go-cue is often followed by inhibition during the outcome. In addition to mPFC

inputs, this inhibition may also be mediated by synaptic inhibition from other local serotonin neurons or from autoinhibition by means of somato-dendritically expressed 5-HT_{1A} receptors. Subsequent analyses will examine uncertainty representations in mPFC. Future experiments could record from identified dorsal-raphe-projecting mPFC neurons as well as dorsal raphe GABA neurons to test these ideas.

The optogenetic activation parameters were selected to mimic observed firing rates of serotonin neurons and it has been previously shown that this type of axon activation increases serotonin in this region (Miyazaki et al., 2020). However, the exact consequences of this manipulation are not certain. Optogenetic activation of serotonin neuron axons may evoke back-propagating action potentials that mediate the forms of serotonin neuron inhibition described above. However, these effects may mimic endogenous patterns of secondary activity. It is also unclear if the observed effects of axon stimulation are mediated by serotonin or glutamate. Many serotonin neurons that express vesicular glutamate transporter 3 target the cortex (Ren et al., 2019; Okaty et al., 2020). The timescales of behavioral (tens of seconds) and neural effects (multiple seconds) would permit contributions from both neurotransmitters. Glutamate and serotonin may also be co-released and effects mediated by both.

Experiments and analyses are ongoing, but thus far the results are consistent with serotonin neuron inputs communicating information about uncertainty to mPFC in order to modulate value representations and thereby facilitate adaptive learning.

4.4 Methods

Animals and surgery. We used 5 male and female mice, backcrossed with C57BL/6J and heterozygous for Cre recombinase under the control of the serotonin transporter gene (Slc6a4$^{tm1(cre)Xz}$, The Jackson Laboratory, 014554; Zhuang et al., 2005). Surgery was performed on mice between the ages of 4–8 weeks, under isoflurane anesthesia (1.0–1.5% in O_2) and in aseptic condtions. During surgery, custom-made titanium headplates were surgically

attached to the skull using dental adhesive (C&B-Metabond, Parkell). After the surgeries, analgesia (ketoprofen, 5 mg kg^{-1} and buprenorphine, 0.05–0.1 mg kg^{-1}) was administered to minimize pain and aid recovery. We implanted a custom microdrive targeting right mPFC with tetrodes wrapped around an optic fiber, entering through a craniotomy at 2.5 mm anterior to bregma and 0.5 mm lateral to midline. A second, fixed optic fiber was implanted targeting left mPFC at 2.5 mm anterior to bregma and −0.5 mm lateral to midline.

For all experiments, mice were given at least one week to recover prior to water restriction. During water restriction, mice had free access to food and were monitored daily in order to maintain 80% of their baseline body weight. All mice were housed in reverse light cycle (12h dark/12h light, dark from 08:00–20:00) and all experiments were conducted during the dark cycle between 10:00 and 18:00. All surgical and experimental procedures were in accordance with the *National Institutes of Health Guide for the Care and Use of Laboratory Animals* and approved by the Johns Hopkins University Animal Care and Use Committee.

Behavioral task. Before training on the dynamic foraging task, water-restricted mice were habituated to head fixation for 1–3 d with free access to water from the provided spouts (two 21 ga stainless steel tubes separated by 4 mm) placed in front of the 38.1 mm acrylic tube in which the mice rested. The spouts were mounted on a micromanipulator (DT12XYZ, Thorlabs) with a custom digital rotary encoder system to reliably determine the position of the lick spouts in XYZ space with 5–10 μm resolution (Bari et al., 2019). Each spout was attached to a solenoid (ROB-11015, Sparkfun) to enable retraction (see Behavioral tasks: dynamic foraging). The odors used for the cues (p-cymene and (−)-carvone) were dissolved in mineral oil at 1:10 dilution (30 μl) and absorbed in filter paper housed in syringe adapters (Whatman, 2.7 μm pore size). The adapters were connected to a custom-made olfactometer (Cohen et al., 2012) that diluted odorized air with filtered air by 1:10 to produce a 1.0 L min^{-1} flow rate. The same flow rate was maintained outside of the cue period so that flow rate was constant throughout the task.

Licks were detected using a custom circuit (Janelia Research Campus 2019-053). Task

events were controlled and recorded using custom code (Arduino) written for a microcontroller (ATmega16U2 or ATmega328). Water rewards were 2–4 μl, adjusted for each mouse to maximize the number of trials completed per session and to keep sessions around 60 minutes. Solenoids (LHDA1233115H, The Lee Co) were calibrated to release the desired volume of water and were mounted on the outside of the dark, sound-attenuated chamber used for behavior tasks. White noise (2–60 kHz, Sweetwater Lynx L22 sound card, Rotel RB-930AX two-channel power amplifier, and Pettersson L60 Ultrasound Speaker), was played inside the chamber to block any ambient noise.

During the 1–3 days of habituation, mice were trained to lick both spouts to receive water. Water delivery was contingent upon a lick to the correct spout at any time. Reward probabilities were chosen from the set $\{0, 1\}$ and reversed every 20 trials.

In the second stage of training (5–12 d), the trial structure with odor presentation was introduced. Each trial began with the 0.5 s delivery of either an odor "go cue" ($P = 0.95$) or an odor "no-go cue" ($P = 0.05$). Following the go cue, mice could lick either the left or the right spout. If a lick was made during a 1.5 s response window, reward was delivered probabilistically from the chosen spout. The unchosen spout was retracted at the time of the tongue contacting the other spout so that mice would not try to sample both spouts within a trial. The unchosen spout was replaced 2.5 s after cue onset. Following a no-go cue, any lick responses were neither rewarded nor punished. Reward probabilities during this stage were chosen from the set $\{0, 1\}$ and reversed every 20–35 trials. During this period of training only, water was occasionally manually delivered to encourage learning of the response window and appropriate switching behavior. Reward probabilities were then changed to $\{0.1, 0.9\}$ for 1–2 days of training prior to introducing the final stage of the task. Rewards were never "baited," as in previous versions of the task (Sugrue et al., 2004; Lau and Glimcher, 2005; Tsutsui et al., 2016; Bari et al., 2019). We did not penalize switching with a "changeover delay." If a directional lick bias was observed in one session, the lick spouts were moved horizontally $50 - -300$ μm in the opposite direction prior to the following session.

After the 1.5 s response window, inter-trial intervals were generated as draws from an exponential distribution with a rate parameter of 0.3 and a maximum of 30 s. This distribution results in a flat hazard rate for inter-trial intervals such that the probability of the next trial did not increase over the duration of the inter-trial interval (Luce, 1986). Inter-trial intervals (go-cue on to go-cue on) were 7.27 ± 3.65 s on average (range 2.5–32.5 s). As in previous studies, mice made a leftward or rightward choice in greater than 99% of trials (Bari et al., 2019). Mice completed 343 ± 71.8 trials per session (range 117–570 trials).

In the final stage of the task, the reward probabilities assigned to each lick spout were drawn pseudorandomly from the set $\{0.1, 0.5, 0.9\}$. The probabilities were assigned to each spout individually with block lengths drawn from a uniform distribution of 20–35 trials. To stagger the blocks of probability assignment for each spout, the block length for one spout in the first block of each session was drawn from a uniform distribution of 6–21 trials. For each spout, probability assignments could not be repeated across consecutive blocks. To maintain task engagement, reward probabilities of 0.1 could not be simultaneously assigned to both spouts. If one spout was assigned a reward probability greater than or equal to the reward probability of the other spout for 3 consecutive blocks, the probability of that spout was set to 0.1 to encourage switching behavior and limit the creation of a direction bias. If a mouse perseverated on a spout with reward probability of 0.1 for 4 consecutive trials, 4 trials were added to the length of both blocks. This procedure was implemented to keep mice from choosing one spout until the reward probability became high again.

The probability of the task generating the special case probability transitions (Figure 4-1b) was enhanced when a new probability was being selected for one of the spouts at the end of a block. At this point, if one of the current probabilities was equal to 0.1 then we forced the special case transitions with $P = 1/3$. Medium ($P = 0.5$) and low ($P = 0.1$) were switched to low and high ($P = 0.9$), or high and low were switched to low and high simultaneously. Forced transitions were not allowed to occur in consecutive probability changes. This design increased the frequency of these transitions by $\sim$ 3x without drastically altering task structure

or reward statistics.

To minimize spontaneous licking, we enforced a 1 s no-lick window prior to odor delivery. Licks within this window were punished with a new randomly-generated inter-trial interval, followed by a 2.5 s no-lick window. Implementing this window significantly reduced spontaneous licking throughout the entirety of behavioral experiments.

Electrophysiology. We recorded extracellular signals from neurons at 32 or 30 kHz using a Digital Lynx 4SX (Neuralynx Inc,) or Intan Technologies RHD2000 system (with RHD2132 headstage), respectively. The recording systems were connected to 8 implanted tetrodes (32–64 channels, nichrome wire, PX000004, Sandvik) fed through 39 ga polyimide guide tubes that could be advanced with the turn of a screw on a custom, 3D-printed microdrive. The impedances of each wire in the tetrodes were reduced to 200–300 kΩ by gold plating. The tetrodes were wrapped around a 200μm optic fiber used for optogenetic identification. After each recording session, the tetrode-optic-fiber bundle was driven down 75μm. The median signal was subtracted from raw recording traces across channels and bandpass-filtered between 0.3–6 kHz using custom MATLAB software. To detect peaks, the bandpass-filterd signal, x, was thresholded at $4\sigma_n$ where σ_n=median($\frac{|x|}{0.6745}$) (Quiroga et al., 2004). Detected peaks were sorted into individual unit clusters offline (Spikesort 3D, Neuralynx Inc.) using waveform energy, peak waveform amplitude, minimum waveform trough, and waveform principal component analysis. We used two metrics of isolation quality as inclusion criteria: L-ratio (< 0.05) (Schmitzer-Torbert et al., 2005) and fraction of interspike interval violations ($< 0.1\%$ interspike intervals < 2 ms).

Targeting of mPFC was confirmed by performing electrolytic lesions of the tissue (20 s of 20 μA direct current across two wires of the same tetrode) and examining the tissue after perfusion.

Viral injections. To express channelrhodopsin-2 (ChR2) in dorsal raphe serotonin neurons, we pressure-injected 810 nl of rAAV5-EF1a-DIO-hChR2(H134R)-EYFP (3×10^{13} GC ml^{-1}) into the dorsal raphe of *Sert*-Cre mice at a rate of 1 nl/s (MMO-220A, Narishige). We made

three injections of 270 nl at the following coordinates: $\{4.63, 4.57, 4.50\}$ mm posterior of bregma, $\{0.00, 0.00, 0.00\}$ mm lateral from the midline, and $\{2.80, 3.00, 3.25\}$ mm ventral to the brain surface. The pipette was inserted through a craniotomy at -5.55 mm posterior to bregma and aligned to midline, using a 16° posterior angle. Before the first injection, the pipette was left at the most ventral coordinate for 10 minutes. After each injection, the pipette was withdrawn 50μm and left in place for 5 min. The craniotomy was covered with silicone elastomer (Kwik-Cast, WPI) and dental cement.

Optogenetic stimulation. Blue light (473 nm wavelength, 15 mW) was delivered using a diode-pumped solid-state laser (Laserglow) and a shutter (Uniblitz). The shutter was open for 10 ms pulses at 50 Hz during the odor go-cue (500 ms). On sham control days, patch cords from the laser were placed inside the protective cone surrounding the implanted microdrive. On stimulation days, the patch cords were connected to the implanted optic fibers with ceramic mating sleeves.

Data analysis. All analyses were performed with MATLAB (Mathworks). All data are presented as mean ± S.D. unless reported otherwise. All statistical tests were two-sided.

Data analysis: generative model of behavior with static learning. We applied a generative RL model of behavior in the foraging task with static learning rates (Daw et al., 2006; Bari et al., 2019). This RL model estimates action values ($Q_l(t)$ and $Q_r(t)$) on each trial to generate choices. Choices are described by a random variable, $c(t)$, corresponding to left or right choice, $c(t) \in \{l, r\}$. The value of a choice is updated as a function of the RPE, and the rate at which this learning occurs is controlled by the learning rate parameter α. Because we observed asymmetric learning from rewards and no rewards (Figure 4-1b), consistent with previous reports (Bari et al., 2019), we included separate learning rates for the different outcomes. For example, if the left spout was chosen, then

$$Q_l(t+1) = \begin{cases} Q_l(t) + \alpha_{(+)}\delta(t), & \text{if } \delta(t) > 0 \\ Q_l(t) + \alpha_{(-)}\delta(t), & \text{if } \delta(t) < 0 \end{cases}$$
$$Q_r(t+1) = \zeta Q_r(t),$$

where $\delta(t) = R(t) - Q_l(t)$ and ζ represents the forgetting rate parameter. The forgetting rate captures the increasing uncertainty about the value of the unchosen spout.

The Q-values are used to generate choice probabilities through a softmax decision function (Daw et al., 2006):

$$P(c(t) = r) = \frac{1}{1 + e^{-\beta(Q_r(t) - Q_l(t) + bias)}},$$
$$P(c(t) = l) = 1 - P(c(t) = r),$$

where β, the "inverse temperature" parameter, controls the steepness of the sigmoidal function. In other words, β controls the level of exploration versus exploitation with respect to the relative action values.

Data analysis: generative model of behavior with meta-learning. We observed mouse behavior that the static learning model failed to capture and that suggested that learning rate was not constant over time. Thus, we added a component to the model that modulates RPE magnitude and $\alpha_{(-)}$ ("meta-learning"). Because learning should be slow in stable but variable environments, expected uncertainty scaled RPEs, such that learning is decreased when expected uncertainty is high. If the left spout was chosen, the values of actions were updated according to

$$Q_l(t+1) = \begin{cases} Q_l(t) + \alpha_{(+)}\delta(t)(1 - \epsilon(t)), & \text{if } \delta(t) > 0 \\ Q_l(t) + \alpha_{(-)}(t)\delta(t)(1 - \epsilon(t)), & \text{if } \delta(t) < 0 \end{cases}$$
$$Q_r(t+1) = \zeta Q_r(t),$$

where ϵ is an evolving estimate of expected uncertainty calculated from the history of unsigned RPEs:

$$\upsilon(t) = |\delta(t)| - \epsilon(t),$$
$$\epsilon(t+1) = \epsilon(t) + \alpha_\upsilon \upsilon(t).$$

The rate of RPE magnitude integration is controlled by α_υ. Deviations from the expected uncertainty are captured by unexpected uncertainty, υ, and may indicate that a change has occurred in the environment. Changes in the environment should drive learning to adapt behavior to new contingencies so $\alpha_{(-)}$ varies as a function of how surprising recent outcomes are:

$$\alpha_{(-)}(t) = \begin{cases} \alpha_{(-)}(t-1) & \text{if } \delta(t) > 0 \\ \psi(\upsilon(t) + \alpha_{(-)_0}) + (1-\psi)(\alpha_{(-)}(t-1)) & \text{if } \delta(t) < 0 \end{cases}$$

where $\alpha_{(-)_0}$ is the baseline learning rate from no reward and ψ controls how quickly unexpected uncertainty is integrated to update $\alpha_{(-)}$. As it is formulated, $\alpha_{(-)}$ increases after surprising no reward outcomes. This learning rate was not allowed to be less than 0, such that

$$\alpha_{(-)}(t) = 0, \text{if } \alpha_{(-)}(t) < 0$$

The Q-values are used to generate choice probabilities through a softmax decision function (Daw et al., 2006):

$$P(c(t) = r) = \frac{1}{1 + e^{-\beta(Q_r(t) - Q_l(t) + bias)}},$$
$$P(c(t) = l) = 1 - P(c(t) = r),$$

where β, the "inverse temperature" parameter, controls the steepness of the sigmoidal function. In other words, β controls the level of exploration versus exploitation with respect to the relative action values.

Data analysis: Model fitting. We fit and assessed models using MATLAB (Mathworks) and the probabilistic programming language, Stan (https://mc-stan.org/) with the MATLAB interface, MatlabStan (https://mc-stan.org/users/interfaces/matlab-stan). Stan was used to construct hierarchical models with mouse-level hyperparameters to govern session-level parameters. For each session, each parameter in the model (for example, α_ϵ for the meta-learning model) was modeled as a draw from a mouse-level distribution with mean μ and variance σ. Models were fit using noninformative (uniform distribution) priors for session-level parameters ($[0, 1]$ for all parameters except β which was $[0, 10]$) and weakly informative ($\mu \sim \mathcal{N}(0, 1)$, $\sigma \sim$ half-Cauchy$(0, 3)$) priors for mouse-level hyperparameters. For some mice with fewer sessions, more informative mouse-level hyperparmeters were used to achieve model convergence under the assumption that individual mice behave similarly across days. This hierarchical construction mitigated the typical variability of point estimates for session-level parameters that results from other methods of estimation. Stan uses full Bayesian statistical inference to generate posterior distributions of parameter estimates using Hamiltonian Markov chain Monte Carlo sampling (Carpenter et al., 2017). The parameters for updating expected uncertainty, α_υ, and for updating the negative RPE learning rate, ψ, were ordered such that $\psi > \alpha_\upsilon$. The ordering operated under the assumption that learning rate should be integrated more quickly to detect change. The ordering also helped models to converge more quickly.

Data analysis: extracting model parameters and variables, behavior simulation. For extracting model variables (like expected uncertainty), we took at least 4,000 draws from the Hamiltonian Markov Chain Monte Carlo samples of session-level parameters, ran the model agent through the task with the actual choices and outcomes, and averaged each model variable across runs. For comparisons of individual parameters across behavioral models, we obtained maximum *a posteriori* parameter values by approximating the mode of the distribution: binning the values in 50 bins and taking the median value of the most populated bin. For simulations of behavior, we took at least 4,000 draws from the Hamiltonian Markov Chain Monte Carlo samples of mouse-level parameters and simulated behavior and outcomes

in a number of random sessions per sample. For the transition analysis, that number was proportional to the number of rare transitions that each animal contributed to the actual data. For other analyses that number was fixed.

Data analysis: linear regression models of neural activity. For comparisons of firing rates to the behavioral-model-generated variables we regressed spike counts in the last 1 s of the inter-trial interval on those variables using the MATLAB function "fitglm" with a Poisson distribution.

Histology. After experiments were completed, mice were euthanized with an overdose of isoflurane, exsanguinated with saline, and perfused with 4% paraformaldehyde. The brains were cut in 100-μm-thick coronal sections and mounted on glass slides. We validated expression of rAAV5-EF1a-DIO-hChR2(H134R)-EYFP epifluorescence images of dorsal raphe and mPFC (Zeiss Axio Zoom.V16). We confirmed targeting of the optic-fiber-tetrode bundle to the mPFC by location of the electrolytic lesion.

Chapter 5

Conclusions, limitations, and future directions

Serotonin neuron function has proven difficult to specify. This may be the case because serotonin neurons do not perform a unitary function or, similarly, that a unifying explanation of their function would be too general to be of much use. Some pioneering theories have proposed computational roles for these neurons, but they have been largely untested. Another limitation has been the relative scarcity of recordings from identified serotonin neurons during awake behavior. The work described in this dissertation was carried out with the intent to make progress in linking action potentials in serotonin neurons to behavior. We designed a dynamic foraging task (originally adapted from a primate task by Bari et al., 2019) for head-restrained mice, in which the statistics of reward varied. These reward dynamics allowed us to compare generative models of behavior that describe meta-learning. These models allowed us to constrain the possible types of cognitive processes that could produce such behavior. The decision making that we observed was consistent with the modulation of learning by uncertainty. Specifically, expected uncertainty could temper learning while unexpected uncertainty augments it. These roles for uncertainty are consistent with normative models of learning and behavioral evidence from humans and other species.

Additional evidence for the idea that the brain performs this type of meta-learning was provided by recordings from individual serotonin neurons in foraging mice. We found that

the activity of roughly half of serotonin neurons correlated with the expected uncertainty variable ($\epsilon(t)$) estimated from behavior over long timescales. Phasic activity supported a potential mechanism, consistent with the formulation of the model, whereby expected uncertainty is updated by unexpected uncertainty ($\upsilon(t)$) at the time of the outcome. These relationships generalized to a Pavlovian behavioral context, in which outcomes were no longer contingent upon the animals' actions. The meta-learning model also makes predictions about behavioral change when uncertainty modulation of learning is removed. Reversible inhibition of serotonin neuron activity revealed changes in decision making behavior consistent with these predictions.

Because the serotonin neurons we recorded displayed considerable heterogeneity in responses to task features and model variables, we examined the possibility that the population projecting to the medial prefrontal cortex conveyed uncertainty. We recorded from single neurons in the medial prefrontal cortex, whose activity has been shown to represent value in a similar foraging task. We confirmed the presence of relative and total value representations with decision variables generated from our meta-learning model. Optogenetic activation of local serotonin neuron axons drove observable changes in adaptive learning behavior. These effects could potentially be explained as an enhancement of expected uncertainty. Findings are preliminary, but may suggest that the stimulation modulated value representations in a manner consistent with this hypothesis.

While recordings of action potentials constitute some ground truth about neuronal activity, our model is limited by the nature of models and our interpretations by methodological constraints. But these findings warrant additional examination and raise further questions about serotonin neuron function, learning, and decision making.

5.1 Limitations and future directions of foraging and modeling

Foraging tasks such as the one we used (also referred to as "restless bandit" tasks) are useful for studying continuous learning and decision making behavior across species. While our particular design choices create a rich set of reward statistics for testing hypotheses about these processes, there are a few limitations. As noted in the Discussion of Chapter 2, certain asymmetries in task structure result in rewards carrying more information about which spout is "good enough" ($P(R) = 0.9$ or $P(R) = 0.5$ as opposed to $P(R) = 0.1$). Deterministic and coupled contingencies ($P(R) = 1$ and $P(R) = 0$) could be implemented to saturate learning from no rewards as has been used in classic reversal tasks. In this case, no reward outcomes provide as much information as rewards. However, in this case reward statistics become very restricted and asymmetries in learning are often still observed (Clarke et al., 2004, 2007). In our task, the asymmetry in information is also a result of binary outcomes, i.e., rewards are delivered or not. Reward volume could be manipulated in order to tease apart if learning from less-than-expected rewards differs from learning from lack of rewards.

One of the reasons foraging tasks have been so widely used is that the behavior they produce is quite amenable to description by computational models. In particular, the evoked behavior can be well-characterized by the same simple reinforcement learning algorithms across multiple species. But all models are false oversimplifications. Despite this fact, they are still crucial tools in understanding potential latent cognitive processes and their neural implementation. In fact, their falsifiability, and thus their ability make testable predictions, is part of what makes them so useful.

Reinforcement learning algorithms (model free), like the ones described in this dissertation, are challenged by the abilities of humans and other species to learn about the structure of the world and make inferences from that knowledge (model based). As such, it may very well be the case that mice have some higher order knowledge about the assigned reward

probabilities, their frequency, block lengths, as well as the fact that reward probabilities cannot be 0.1 at both spouts simultaneously. This knowledge could be used by the mice to make inferences about the value of the unchosen spout or the likelihood that the reward probability assignment has changed. Our model does not take into account these possibilities, although it still does a reasonable job at characterizing behavior. One reason for this success may be that the model approximates a more sophisticated process of inference in this specific behavioral context. Computation of unexpected uncertainty, for example, can be seen as a type of change detection mechanism. Bayesian inference models that compute change point probability or variants that assume knowledge of task structure could be fit to these data to see if they make the same predictions as the meta-learning model (Nassar et al., 2012; McGuire et al., 2014).

Model-based and model-free models of behavior are not mutually exclusive. Learning and decision making behavior has been described as being a product of both types of processes. Raised in the Discussion section of Chapter 3, one idea is that both are present in this foraging behavior and their relative contribution may be mediated by serotonin (Iigaya et al., 2018). This idea is consistent with our findings, but other tasks would be better suited to examine it. The two-step task, for example, involves probabilistic state transitions between multistep decisions (Daw et al., 2011). This design allows for the differentiation between model-free and model-based components of learning.

The meta-learning model is also limited by the use of binary outcomes described above. As a consequence of this design choice, there is some correlation between expected uncertainty and reward rate. When reward rate is high ($P = 0.9$), expected uncertainty is low, for example. Again, varying reward size could provide useful insight. By changing reward volumes, the task can provide reward dynamics with equal reward rate (in terms of volume per trial) but different amounts of uncertainty (variance in reward volumes). That being said, models using reward rate to drive learning or exploration failed to explain behavior. Additionally, the results of chemogenetic inhibition experiments are consistent with serotonin

neuron signaling uncertainty as opposed to reward rate. However, model space is infinite and there may be formulations of models that appropriately capture learning as a function of reward rate.

In addition to limits regarding task design and models of the behavior it produces, these experiments do not address many of the other functions with which serotonin neuron activity has been associated. Some of these functions, however, may be related to our findings. Previous studies have related serotonin neuron activity to the value of outcomes, both rewarding and aversive (Miyazaki et al., 2011; Cohen et al., 2015; Li et al., 2016; Matias et al., 2017). The particulars of these findings suggest that serotonin neuron activity may connote state value. It is possible that this information could be used to drive learning in a similar way as uncertainty, but not in the manner proposed by the opponency or global reward state models that we tested. To explain foraging behavior, there would need to be some change detection component to the state value computation in order to drive learning adaptively.

One potential explanation of the relationship of serotonin neuron activity to value comes from the observation that the expected uncertainty computation is one that generalizes across actions. As we suggested, this generalized uncertainty may refer to a behavioral policy or context. If this is the case, the response to a cue that predicts a smaller reward (as opposed to a separate cue predicting a larger reward in the same task), may have smaller expected uncertainty due to smaller reward prediction errors. Reward prediction errors, of course, would have to be present even though the outcome is fully predicted by the cue. Indeed, persistent responses to fully-predicted outcomes have been shown in mouse dopamine neurons (Cohen et al., 2012). These responses may reflect a distributional encoding of reward prediction errors wherein a spectrum of pessimistic and optimistic dopamine neurons have differing value expectations and thus, different reward prediction error magnitudes (Dabney et al., 2020). Consequently, serotonin neurons may signal uncertainty in a distributional fashion as well. However, the persistent dopamine neuron responses during the outcome may not be related to the reward prediction error (Schultz et al., 2017).

One theory proposed that serotonin is involved in the temporal discounting of the value of future states (Doya et al., 2002). Experiments in which activation of serotonin activation enhances waiting or persistence for reward have been explained in this framework (Miyazaki et al., 2011, 2014; Fonseca et al., 2015; Lottem et al., 2018). Our task can be viewed as a single state task, in which an action only results in the immediate outcome and does not affect later outcomes. Given this design, we cannot explicitly test that theory. However, enhanced waiting and persistence may also be explained as modulation of how the brain learns from outcomes or lack thereof.

Low blood contents of serotonin and the relative success of selective serotonin reuptake inhibitors observed in humans with major depression, along with other observations, have led some to posit a link between the neuromodulator, mood, and affective state (Mann et al., 1992; Fournier et al., 2010). Without a mechanistic understanding of how the serotonin system is modified in both the disorder and by the drug, it is difficult to draw strong conclusions about the work presented here in the context of major depression. That being said, there are some interesting parallels. Some studies have shown that subjects with depression have an increased sensitivity to negative information in terms of learning and neural responses (Hamilton and Gotlib, 2008; Dombrovski et al., 2015; Segarra et al., 2016; Xie et al., 2020) as well as reduced neural and behavioral responsivity to rewards (Vrieze et al., 2013; Dombrovski et al., 2013; Brown et al., 2020). One of these studies showed an insensitivity to the change in contingency between actions and outcomes (Dombrovski et al., 2013). Another showed disrupted value signals in the ventromedial prefrontal cortex that correlated with a disruption in outcome driven choice behavior (Brown et al., 2020).

A couple groups have theorized specific interactions between mood and learning (Rutledge et al., 2014; Eldar and Niv, 2015; Eldar et al., 2016). In one of the formulations, mood is modeled as a history of recent prediction errors that modulate the perception of the value of outcomes (Eldar and Niv, 2015). The model quantified the effect of mood on outcome perception and showed that it is enhanced in subjects with mood instability. The model also

predicted how this enhancement drives oscillations in mood and learning in those subjects. In a separate study, a regression model showed that subjective reports of happiness could be predicted from expected values and reward prediction errors (Rutledge et al., 2014). Supported by these notions, mood may be a way to form generalized expectations about reward statistics to tune how outcomes drive behavior (Eldar et al., 2016).

Given the relationship we observed between serotonin neuron activity and learning from less-than-expected outcomes, some of these symptoms of depression and effects of mood may be consistent with alterations in serotonin system function. A better understanding of serotonin system dysfunction in depression is necessary to make this connection. Further, a more thorough and specific understanding of learning differences in subjects with depression should be achieved through behavior and computational modeling. If findings are consistent, such an approach could even be leveraged as a diagnostic tool.

5.2 Limitations and future directions of serotonin neuron identification and sampling

Methodological constraints in the extracellular recordings of action potentials are well-known, but worth considering. These types of recordings are biased towards neurons with a certain morphology and distribution of conductances. These biases can be affected by the position of the electrode relative to the neuron and its processes, so anatomy relative to angle of penetration matters as well. A technically-challenging and low-throughput alternative that circumvents these biases is to use *in vivo* whole cell or cell-attached recording configurations (Aghajanian et al., 1978; Aghajanian and Vandermaelen, 1982; Hajós and Sharp, 1996; Schweimer and Ungless, 2010; Schweimer et al., 2011).

In our recordings from dorsal raphe, our implantation coordinates biased sampling to neurons along the midline. Given the angle we used (16° posterior to bregma), we likely sampled dorsally in the middle of the rostro-caudal extent and ventrally in the more rostral portions. This path was chosen to maximize the volume of dorsal raphe sampled in a single

animal, but certainly biases the recordings towards certain subpopulations—or at least away from those in the lateral wings. Sampling that is more topographically comprehensive will be necessary for understanding the function of all subpopulations of serotonin neurons.

The criteria we used for optogenetic identification of serotonin neurons were relatively strict to avoid false positives at the expense of false negatives. Most of the neurons in dorsal raphe are serotonin neurons, but only a small percentage of the neurons we recorded responded to laser stimulation with enough reliability and short enough latency to be included in our analyses. This approach yields high confidence in the identity of included neurons, but certain electrophysiological properties or tropisms in virus expression across subpopulations may result in sampling biases. Future experiments could target virus expression to certain subpopulations to avert some of these biases.

5.3 Limitations and future directions of neural activity analyses

In serotonin and medial prefrontal cortex neurons we observed phasic activity dynamics during the trial period. In serotonin neurons, activity following cue onset correlated with expected uncertainty while activity during the outcome correlated with violations in that expectation, or unexpected uncertainty. The temporal proximity of these two epochs makes it somewhat difficult to distinguish the relationship between activity, the cue, motor behavior, and outcome. The timing of the peak of expectation-related activity correlated with lick latency, preceding the tongue touching the spout by a couple hundred milliseconds. However, it is still unclear if this response is related to detection of the cue or initiation of the lick. In the Pavlovian version of the task we saw responses to the cue that preceded the first anticipatory lick, but the question still remains. Future tasks could implement a delay period between cue and choice or outcome during which the animal is required to withhold their choice behavior. Combined with high-speed video of licks and other motor behavior, one could begin to tease apart how uncertainty-related serotonin neuron activity is aligned to

predictive cues and actions. It is possible that the activity is aligned to whatever is the most reliable, temporal predictor of the outcome. In the Pavlovian task, this would be the cue, but in the foraging task this may include both the cue and the action.

Results from and analysis of medial prefrontal cortex recordings are ongoing. In the future, the phasic activity will be analyzed in relationship to actions, outcomes, and potential cognitive variables. As mentioned in the Discussion section of Chapter 4, we will also look for representations of uncertainty in this region, since it contributes substantial inputs to dorsal raphe. Further analyses will also compare the activity of identified dorsal raphe serotonin neurons to our larger database of unidentified neurons in the same region.

5.4 Limitations and future directions of serotonin neuron activity manipulation

Exogenous manipulations of neural activity have long-been employed to test causality between endogenous activity and behavioral function. While tools and techniques for manipulation have become more precise, they are still limited. The inhibitory DREADDs receptor used in our experiments, hM4Di, has been shown to effectively suppress serotonin neuron excitability (Armbruster et al., 2007; Ray et al., 2011; Teissier et al., 2015). Whole-cell recordings *in vitro* showed that while reduction in excitability was about 40% on average, there was considerable heterogeneity across individual neurons in the magnitude of inhibition (Ray et al., 2011). One population that is effectively inhibited appears to be those neurons that innervate medial prefrontal cortex. DREADDs-mediated inhibition of serotonin neurons *in vivo* resulted in a large decrease in medial prefrontal cortex serotonin contents measured with microdialysis in anesthetized mice (Teissier et al., 2015). Similar to increasing specificity in observations of activity, confirming efficacy and subsequently targeting subpopulations of neurons with this sort of manipulation will be an effective strategy in understanding serotonin neuron function.

In the context of our experiments, these previous findings have a few implications. We observed heterogeneity in responses of serotonin neurons. Relationships to uncertainty in

particular were conveyed with both positive and negative correlations. Inhibiting serotonin neurons *en masse* resulted in behavioral changes consistent with the removal of uncertainty modulation of learning. This manipulation could be interpreted as taking these forms of uncertainty offline. In this way, baseline firing rates of neurons negatively correlated with uncertainty may be necessary for signaling high uncertainty, but may be silenced by the manipulation. A related possibility is that these negative correlations are a postsynaptic consequence of the activity of positively correlated neurons. This possibility could be enabled by local connectivity and expression of inhibitory 5-HT_{1A} receptors. Separately, the manipulation may be more effective at inhibiting positively correlated neurons or inhibiting certain modes of firing. Finally, uncertainty signals might be used by disparate postsynaptic targets for distinct behavioral functions. This manipulation then, may selectively inhibit pathways involved in this foraging behavior. These possibilities are highly speculative and require further examination to demonstrate.

Optogenetics allows for more temporally precise manipulation of neural activity, but how these evoked responses relate to endogenous activity is not entirely clear. Optogenetic activation of serotonin neuron axons in medial prefrontal cortex is sufficient to increase serotonin contents in that region (Miyazaki et al., 2020). And when we optogenetically activated serotonin neuron axons in medial prefrontal cortex, we chose stimulation parameters derived from the activity of observed serotonin neurons. Most serotonin neurons were shown to be phasically activated during the cue period and exhibited brief bursts in firing rates as high as 30 Hz. Our tetrodes were also implanted in regions of the dorsal raphe with the highest densities of cortex-projecting neurons. Even so, we do not know which of these neurons, if any, send axonal projections to medial prefrontal cortex.

We chose a stimulation frequency of 50 Hz during the cue to maximize firing rates prior to the receipt of the outcome. In addition to not knowing the typical firing rates of neurons that project to medial prefrontal cortex, it is unclear if stimulation of axons at a certain frequency mimics the effects of action potentials at that frequency. It is also unclear if serotonin neuron

activity in this region is coordinated, so activating large swaths of axons is unlikely to mimic endogenous activity. Another possibility is that this pattern of stimulation elicits release of neurotransmitters that endogenous activity during this epoch would not (Svensson et al., 2018). In these ways, results from gain-of-function experiments can be difficult to interpret. Methods for local inhibition of serotonin neuron axons would avoid some of these issues. Observation of endogenous serotonin neuron activity in medial prefrontal cortex (e.g., imaging of fluorescent reporters of presynaptic calcium activity or postsynaptic ligand binding) could also clarify local serotonin neuron function.

5.5 Conclusion

Behavioral, modeling, and technical limitations are an inevitable burden of experimentation. Only through convergent evidence from many, carefully designed experiments can neuroscience research inch towards a meaningful understanding of the brain. Here, we sought to contribute one small piece to the puzzle that is serotonin neuron function. Foraging behavior, recordings from identified dorsal raphe serotonin neurons and medial prefrontal cortex neurons, manipulations of serotonin neuron activity, and computational modeling support the hypothesis that serotonin neurons modulate learning rate through uncertainty.

Bibliography

Adhikari A, Topiwala MA, Gordon JA. Synchronized activity between the ventral hippocampus and the medial prefrontal cortex during anxiety. *Neuron* 65: 257–269, 2010.

Adhikari A, Topiwala MA, Gordon JA. Single units in the medial prefrontal cortex with anxiety-related firing patterns are preferentially influenced by ventral hippocampal activity. *Neuron* 71: 898–910, 2011.

Aghajanian GK, Lakoski JM. Hyperpolarization of serotonergic neurons by serotonin and LSD: studies in brain slices showing increased K+ conductance. *Brain Research* 305: 181–185, 1984.

Aghajanian GK, Vandermaelen CP. Intracellular recordings from serotonergic dorsal raphe neurons: pacemaker potentials and the effect of LSD. *Brain Research* 238: 463–469, 1982.

Aghajanian GK, Wang RY, Baraban J. Serotonergic and non-serotonergic neurons of the dorsal raphe: reciprocal changes in firing induced by peripheral nerve stimulation. *Brain Research* 153: 169–175, 1978.

Aitken AR, Törk I. Early development of serotonin-containing neurons and pathways as seen in wholemount preparations of the fetal rat brain. *Journal of Comparative Neurology* 274: 32–47, 1988.

Allers K, Sharp T. Neurochemical and anatomical identification of fast- and slow-firing neurones in the rat dorsal raphe nucleus using juxtacellular labelling methods in vivo. *Neuroscience* 122: 193–204, 2003.

Alonso A, Merchán P, Sandoval JE, Sánchez-Arrones L, Garcia-Cazorla A, Artuch R, Ferrán JL, Martínez-de-la Torre M, Puelles L. Development of the serotonergic cells in murine raphe nuclei and their relations with rhombomeric domains. *Brain Structure and Function* 218: 1229–1277, 2013.

Alsiö J, Lehmann O, McKenzie C, Theobald DE, Searle L, Xia J, Dalley JW, Robbins TW. Serotonergic Innervations of the Orbitofrontal and Medial-prefrontal Cortices are Differentially Involved in Visual Discrimination and Reversal Learning in Rats. *Cerebral Cortex* 31: 1090–1105, 2021.

Andrade R, Haj-Dahmane S. Chapter 15 - Cellular effects of serotonin in the CNS. In *Handbook of the Behavioral Neurobiology of Serotonin*, pp. 279 – 288. 2020.

Andrade R, Huereca D, Lyons JG, Andrade EM, McGregor KM. 5-HT1A Receptor-Mediated Autoinhibition and the Control of Serotonergic Cell Firing. *ACS Chemical Neuroscience* 6: 1110–1115, 2015.

Andrade R, Nicoll RA. Pharmacologically distinct actions of serotonin on single pyramidal neurones of the rat hippocampus recorded in vitro. *The Journal of Physiology* 394: 99–124, 1987.

Andrews PW, Bharwani A, Lee KR, Fox M, Thomson JA. Is serotonin an upper or a downer? The evolution of the serotonergic system and its role in depression and the antidepressant response. *Neuroscience and Biobehavioral Reviews* 51: 164–188, 2015.

Aoki M. *State space modeling of time series*. Springer, 1987.

Araneda R, Andrade R. 5-Hydroxytryptamine2 and 5-hydroxytryptamine 1A receptors mediate opposing responses on membrane excitability in rat association cortex. *Neuroscience* 40: 399–412, 1991.

Armbruster BN, Xiang L, Pausch MH, Herlitze S, Roth BL. Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand. *Proceedings of the National Academy of Sciences of the United States of America* 104 (12): 5163–5168, 2007.

Ashby CR, Edwards E, Wang RY. Electrophysiological evidence for a functional interaction between 5-HT1A and 5-HT2A receptors in the rat medial prefrontal cortex: an iontophoretic study. *Synapse* 17: 173–181, 1994.

Athilingam JC, Ben-Shalom R, Keeshen CM, Sohal VS, Bender KJ. Serotonin enhances excitability and gamma frequency temporal integration in mouse prefrontal fast-spiking interneurons. *eLife* 6:e31991, 2017.

Avesar D, Gulledge AT. Selective serotonergic excitation of callosal projection neurons. *Frontiers in Neural Circuits* 6: 12, 2012.

Avesar D, Stephens EK, Gulledge AT. Serotonergic Regulation of Corticoamygdalar Neurons in the Mouse Prelimbic Cortex. *Frontiers in Neural Circuits* 12, 2018.

Awasthi J, Tamada K, Overton E, Takumi T. Comprehensive Topographical Map of the Serotonergic Fibers in the Mouse Brain. *bioRxiv* , 2020.

Azimi Z, Barzan R, Spoida K, Surdin T, Wollenweber P, Mark MD, Herlitze S, Jancke D. Separable gain control of ongoing and evoked activity in the visual cortex by serotonergic input. *eLife* 9: e53552, 2020.

Azmitia EC. Serotonin neurons, neuroplasticity, and homeostasis of neural tissue. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 21: 33S–45S, 1999.

Azmitia EC. Modern views on an ancient chemical: serotonin effects on cell proliferation, maturation, and apoptosis. *Brain Research Bulletin* 56: 413–424, 2001.

Azmitia EC. Serotonin and brain: evolution, neuroplasticity, and homeostasis. *International Review of Neurobiology* 77: 31–56, 2007.

Bach DR, Dolan RJ. Knowing how much you don't know: a neural organization of uncertainty estimates. *Nature Reviews Neuroscience* 13: 572–586, 2012.

Baker KG, Halliday GM, Hornung JP, Geffen LB, Cotton RG, Törk I. Distribution, morphology and number of monoamine-synthesizing and substance P-containing neurons in the human dorsal raphe nucleus. *Neuroscience* 42: 757–775, 1991.

Baker KG, Halliday GM, Törk I. Cytoarchitecture of the human dorsal raphe nucleus. *The Journal of Comparative Neurology* 301: 147–161, 1990.

Balleine BW, Dickinson A. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37: 407–419, 1998.

Bang SJ, Jensen P, Dymecki SM, Commons KG. Projections and interconnections of genetically defined serotonin neurons in mice. *European Journal of Neuroscience* 35: 85–96, 2012.

Bari A, Theobald DE, Caprioli D, Mar AC, Aidoo-Micah A, Dalley JW, Robbins TW. Serotonin modulates sensitivity to reward and negative feedback in a probabilistic reversal learning task in rats. *Neuropsychopharmacology* 35: 1290–1301, 2010.

Bari BA, Grossman CD, Lubin EE, Rajagopalan AE, Cressy JI, Cohen JY. Stable representations of decision variables for flexible behavior. *Neuron* 103: 922–933, 2019.

Barnes NM, Ahern GP, Becamel C, Bockaert J, Camilleri M, Chaumont-Dubel S, Claeysen S, Cunningham KA, Fone KC, Gershon M, Di Giovanni G, Goodfellow NM, Halberstadt AL, Hartley RM, Hassaine G, Herrick-Davis K, Hovius R, Lacivita E, Lambe EK, Leopoldo M, Levy FO, Lummis SCR, Marin P, Maroteaux L, McCreary AC, Nelson DL, Neumaier JF, Newman-Tancredi A, Nury H, Roberts A, Roth BL, Roumier A, Sanger GJ, Teitler

M, Sharp T, Villalón CM, Vogel H, Watts SW, Hoyer D. International Union of Basic and Clinical Pharmacology. CX. Classification of Receptors for 5-hydroxytryptamine; Pharmacology and Function. *Pharmacological Reviews* 73: 310–520, 2021.

Barre A, Berthoux C, De Bundel D, Valjent E, Bockaert J, Marin P, Bécamel C. Presynaptic serotonin 2A receptors modulate thalamocortical plasticity and associative learning. *Proceedings of the National Academy of Sciences of the United States of America* 113: E1382–E1391, 2016.

Baumgartner HR, Born GV. Effects of 5-hydroxytryptamine on platelet aggregation. *Nature* 218: 137–141, 1968.

Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS. Learning the value of information in an uncertain world. *Nature Neuroscience* 10: 1214–1221, 2007.

Bengtson CP, Lee DJ, Osborne PB. Opposing electrophysiological actions of 5-HT on noncholinergic and cholinergic neurons in the rat ventral pallidum in vitro. *Journal of Neurophysiology* 92: 433–443, 2004.

Berthoux C, Barre A, Bockaert J, Marin P, Bécamel C. Sustained Activation of Postsynaptic 5-HT2A Receptors Gates Plasticity at Prefrontal Cortex Synapses. *Cerebral Cortex* 29: 1659–1669, 2019.

Bertsekas DP, Tsitsiklis JN. *Neuro-Dynamic Programming.* Athena Scientific Belmont, 1996.

Biggio G, Fadda F, Fanni P, Tagliamonte A, Gessa GL. Rapid depletion of serum tryptophan, brain tryptophan, serotonin and 5-hydroxyindoleacetic acid by a tryptophan-free diet. *Life Sciences* 14: 1321–1329, 1974.

Blundell JE. Is there a role for serotonin (5-hydroxytryptamine) in feeding? *International Journal of Obesity* 1: 15–42, 1977.

Bocchio M, McHugh SB, Bannerman DM, Sharp T, Capogna M. Serotonin, Amygdala and Fear: Assembling the Puzzle. *Frontiers in Neural Circuits* 10: 24, 2016.

Boulougouris V, Robbins TW. Enhancement of spatial reversal learning by 5-HT2C receptor antagonism is neuroanatomically specific. *Journal of Neuroscience* 30: 930–938, 2010.

Boyden ES, Zhang F, Bamberg E, Nagel G, Deisseroth K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neuroscience* 8: 1263–1268, 2005.

Brigman JL, Mathur P, Harvey-White J, Izquierdo A, Saksida LM, Bussey TJ, Fox S, Deneris E, Murphy DL, Holmes A. Pharmacological or genetic inactivation of the serotonin transporter improves reversal learning in mice. *Cerebral Cortex* 20: 1955–1963, 2010.

Bromberg-Martin ES, Hikosaka O, Nakamura K. Coding of task reward value in the dorsal raphe nucleus. *Journal of Neuroscience* 30: 6262–6272, 2010.

Brown P, Molliver ME. Dual serotonin (5-HT) projections to the nucleus accumbens core and shell: relation of the 5-HT transporter to amphetamine-induced neurotoxicity. *Journal of Neuroscience* 20: 1952–1963, 2000.

Brown RE, Sergeeva OA, Eriksson KS, Haas HL. Convergent excitation of dorsal raphe serotonin neurons by multiple arousal systems (orexin/hypocretin, histamine and noradrenaline). *Journal of Neuroscience* 22: 8850–8859, 2002.

Brown VM, Wilson J, Hallquist MN, Szanto K, Dombrovski AY. Ventromedial prefrontal value signals and functional connectivity during decision-making in suicidal behavior and impulsivity. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology* 45: 1034–1041, 2020.

Burghardt NS, Bush DEA, McEwen BS, LeDoux JE. Acute selective serotonin reuptake inhibitors increase conditioned fear expression: blockade with a 5-HT(2C) receptor antagonist. *Biological Psychiatry* 62: 1111–1118, 2007.

Burghardt NS, Sullivan GM, McEwen BS, Gorman JM, LeDoux JE. The selective serotonin reuptake inhibitor citalopram increases fear after acute treatment but reduces fear with chronic treatment: a comparison with tianeptine. *Biological Psychiatry* 55: 1171–1178, 2004.

Bécamel C, Gavarini S, Chanrion B, Alonso G, Galéotti N, Dumuis A, Bockaert J, Marin P. The serotonin 5-HT2A and 5-HT2C receptors interact with specific sets of PDZ proteins. *The Journal of Biological Chemistry* 279: 20257–20266, 2004.

Cai X, Kim S, Lee D. Heterogeneous coding of temporally discounted values in the dorsal and ventral striatum during intertemporal choice. *Neuron* 69: 170–182, 2011.

Cardin JA, Carlén M, Meletis K, Knoblich U, Zhang F, Deisseroth K, Tsai LH, Moore CI. Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* 459: 663–667, 2009.

Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A. Stan: A Probabilistic Programming Language. *Journal of Statistical Software* 76: 1–32, 2017.

Cazettes F, Reato D, Morais JP, Renart A, Mainen ZF. Phasic Activation of Dorsal Raphe Serotonergic Neurons Increases Pupil Size. *Current Biology* 31: R32–R34, 2021.

Challis C, Beck SG, Berton O. Optogenetic modulation of descending prefrontocortical inputs to the dorsal raphe bidirectionally bias socioaffective choices after social defeat. *Frontiers in Behavioral Neuroscience* 8: 43, 2014.

Chen X, Choo H, Huang XP, Yang X, Stone O, Roth BL, Jin J. The First Structure–Activity Relationship Studies for Designer Receptors Exclusively Activated by Designer Drugs. *ACS Chemical Neuroscience* 6: 476–484, 2015.

Chowdhury S, Yamanaka A. Optogenetic activation of serotonergic terminals facilitates GABAergic inhibitory input to orexin/hypocretin neurons. *Scientific Reports* 6: 36039, 2016.

Clarke HF, Dalley JW, Crofts HS, Robbins TW, Roberts AC. Cognitive inflexibility after prefrontal serotonin depletion. *Science* 304: 878–880, 2004.

Clarke HF, Walker SC, Dalley JW, Robbins TW, Roberts AC. Cognitive inflexibility after prefrontal serotonin depletion is behaviorally and neurochemically specific. *Cerebral Cortex* 17: 18–27, 2007.

Cohen JY, Amoroso MW, Uchida N. Serotonergic neurons signal reward and punishment on multiple timescales. *eLife* 4, 2015.

Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* 482: 85–88, 2012.

Cools R, Roberts AC, Robbins TW. Serotoninergic regulation of emotional and behavioural control processes. *Trends in Cognitive Sciences* 12: 31–40, 2008.

Correia PA, Lottem E, Banerjee D, Machado AS, Carey MR, Mainen ZF. Transient inhibition and long-term facilitation of locomotion by phasic optogenetic activation of serotonin neurons. *eLife* 6:e20975, 2017.

Crockett MJ, Clark L, Tabibnia G, Lieberman MD, Robbins TW. Serotonin Modulates Behavioral Reactions to Unfairness. *Science* 320: 1739–1739, 2008.

Dabney W, Kurth-Nelson Z, Uchida N, Starkweather CK, Hassabis D, Munos R, Botvinick M. A distributional code for value in dopamine-based reinforcement learning. *Nature* 577: 671–675, 2020.

Dahlström A, Fuxe K. Evidence for the existence of monoamine-containing neurons in the

central nervous system. I. Demonstration of monoamines in the cell bodies of brain stem neurons. *Acta Physiologica Scandinavica Supplementum* , 1964.

Dale E, Grunnet M, Pehrson AL, Frederiksen K, Larsen PH, Nielsen J, Stensbøl TB, Ebert B, Yin H, Lu D, Liu H, Jensen TN, Yang CR, Sanchez C. The multimodal antidepressant vortioxetine may facilitate pyramidal cell firing by inhibition of 5-HT3 receptor expressing interneurons: An in vitro study in rat hippocampus slices. *Brain Research* 1689: 1–11, 2018.

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69: 1204–1215, 2011.

Daw ND, Kakade S, Dayan P. Opponent interactions between serotonin and dopamine. *Neural Networks* 15: 603–616, 2002.

Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience* 8: 1704–1711, 2005.

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. Cortical substrates for exploratory decisions in humans. *Nature* 441: 876–879, 2006.

Dayan P, Kakade S, Montague PR. Learning and selective attention. *Nature Neuroscience* 3: 1218–1223, 2000.

De-Miguel FF, Leon-Pinzon C, Noguez P, Mendez B. Serotonin release from the neuronal cell body and its long-lasting effects on the nervous system. *Philosophical Transactions of the Royal Society B* 370, 2015.

Derkach V, Surprenant A, North RA. 5-HT3 receptors are membrane ion channels. *Nature* 339: 706–709, 1989.

Descarries L, Beaudet A, Watkins KC. Serotonin nerve terminals in adult rat neocortex. *Brain Research* 100: 563–588, 1975.

Diederen KM, Schultz W. Scaling prediction errors to reward variability benefits error-driven learning in humans. *Journal of Neurophysiology* 114: 1628–1640, 2015.

Dombrovski AY, Szanto K, Clark L, Aizenstein HJ, Chase HW, Reynolds CF, Siegle GJ. Corticostriatothalamic reward prediction error signals and executive control in late-life depression. *Psychological Medicine* 45: 1413–1424, 2015.

Dombrovski AY, Szanto K, Clark L, Reynolds CF, Siegle GJ. Reward signals, attempted suicide, and impulsivity in late-life depression. *Journal of American Medical Association Psychiatry* 70: 1, 2013.

Dorfman HM, Bhui R, Hughes BL, Gershman SJ. Causal Inference About Good and Bad Outcomes. *Psychological Science* 30: 516–525, 2019.

Doya K. Metalearning and neuromodulation. *Neural Networks* 15: 495–506, 2002.

Doya K, Samejima K, Katagiri Ki, Kawato M. Multiple model-based reinforcement learning. *Neural Computation* 14: 1347–1369, 2002.

Dugué GP, Lörincz ML, Lottem E, Audero E, Matias S, Correia PA, Léna C, Mainen ZF. Optogenetic recruitment of dorsal raphe serotonergic neurons acutely decreases mechanosensory responsivity in behaving mice. *PloS one* 9: e105941, 2014.

Dölen G, Darvishzadeh A, Huang KW, Malenka RC. Social reward requires coordinated activity of nucleus accumbens oxytocin and serotonin. *Nature* 501: 179–184, 2013.

Edelman GM, Gally JA. Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences of the United States of America* 98: 13763–13768, 2001.

Elber-Dorozko L, Loewenstein Y. Striatal action-value neurons reconsidered. *eLife* 7:e34248, 2018.

Eldar E, Niv Y. Interaction between emotional state and learning underlies mood instability. *Nature Communications* 6: 6149, 2015.

Eldar E, Rutledge RB, Dolan RJ, Niv Y. Mood as Representation of Momentum. *Trends in Cognitive Sciences* 20: 15–24, 2016.

Eshel N, Bukwich M, Rao V, Hemmelder V, Tian J, Uchida N. Arithmetic and local circuitry underlying dopamine prediction errors. *Nature* 525: 243–246, 2015.

Faraji M, Preuschoff K, Gerstner W. Balancing new against old information: the role of puzzlement surprise in learning. *Neural Computation* 30: 34–83, 2018.

Feldberg W, Myers RD. A new concept of temperature regulation by amines in the hypothalamus. *Nature* 200: 1325, 1963.

Fernandez SP, Cauli B, Cabezas C, Muzerelle A, Poncer JC, Gaspar P. Multiscale single-cell analysis reveals unique phenotypes of raphe 5-HT neurons projecting to the forebrain. *Brain Structure and Function* 221: 4007–4025, 2016.

Fiorillo CD, Tobler PN, Schultz W. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299: 1898–1902, 2003.

Fletcher PJ, Korth KM. Activation of 5-HT1B receptors in the nucleus accumbens reduces amphetamine-induced enhancement of responding for conditioned reward. *Psychopharmacology* 142: 165–174, 1999.

Fletcher PJ, Korth KM, Chambers JW. Selective destruction of brain serotonin neurons by 5,7-dihydroxytryptamine increases responding for a conditioned reward. *Psychopharmacology* 147: 291–299, 1999.

Fletcher PJ, Ming ZH, Higgins GA. Conditioned place preference induced by microinjection of 8-OH-DPAT into the dorsal or median raphe nucleus. *Psychopharmacology* 113: 31–36, 1993.

Fletcher PJ, Tampakeras M, Yeomans JS. Median raphe injections of 8-OH-DPAT lower

frequency thresholds for lateral hypothalamic self-stimulation. *Pharmacology, Biochemistry, and Behavior* 52: 65–71, 1995.

Fonseca MS, Murakami M, Mainen ZF. Activation of dorsal raphe serotonergic neurons promotes waiting but is not reinforcing. *Current Biology* 25: 306–315, 2015.

Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, Fawcett J. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *Journal of American Medical Association* 303: 47–53, 2010.

Gantz SC, Moussawi K, Hake HS. Delta glutamate receptor conductance drives excitation of mouse dorsal raphe neurons. *eLife* 9:e56054, 2020.

Garattini S, L V. *Serotonin*, vol. 41. Elsevier Publishing Company, 1965.

Geddes SD, Assadzada S, Lemelin D, Sokolovski A, Bergeron R, Haj-Dahmane S, Béïque JC. Target-specific modulation of the descending prefrontal cortex inputs to the dorsal raphe nucleus by cannabinoids. *Proceedings of the National Academy of Sciences of the United States of America* 113: 5429–5434, 2016.

Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis* 1: 515–534, 2006.

Gershman SJ. Empirical priors for reinforcement learning models. *Journal of Mathematical Psychology* 71: 1–6, 2016.

Godlewska BR, Browning M, Norbury R, Cowen PJ, Harmer CJ. Early changes in emotional processing as a marker of clinical response to SSRI treatment in depression. *Translational Psychiatry* 6: e957, 2016.

Haj-Dahmane S, Hamon M, Lanfumey L. K+ channel and 5-hydroxytryptamine1A autoreceptor interactions in the rat dorsal raphe nucleus: Anin vitro electrophysiological study. *Neuroscience* 41: 495–505, 1991.

Hajós M, Sharp T. Burst-firing activity of presumed 5-HT neurones of the rat dorsal raphe nucleus: electrophysiological analysis by antidromic stimulation. *Brain Research* 740: 162–168, 1996.

Hale MW, Hay-Schmidt A, Mikkelsen JD, Poulsen B, Shekhar A, Lowry CA. Exposure to an open-field arena increases c-Fos expression in a distributed anxiety-related system projecting to the basolateral amygdaloid complex. *Neuroscience* 155: 659–672, 2008.

Hale MW, Lowry CA. Functional topography of midbrain and pontine serotonergic systems: implications for synaptic regulation of serotonergic circuits. *Psychopharmacology* 213: 243–264, 2011.

Halliday GM, Li YW, Joh TH, Cotton RG, Howe PR, Geffen LB, Blessing WW. Distribution of monoamine-synthesizing neurons in the human medulla oblongata. *The Journal of Comparative Neurology* 273: 301–317, 1988.

Hamilton JP, Gotlib IH. Neural substrates of increased memory sensitivity for negative stimuli in major depression. *Biological Psychiatry* 63: 1155–1162, 2008.

Hangya B, Ranade S, Lorenc M, Kepecs A. Central Cholinergic Neurons Are Rapidly Recruited by Reinforcement Feedback. *Cell* 162: 1155–1168, 2015.

Harmer CJ, Shelley NC, Cowen PJ, Goodwin GM. Increased positive versus negative affective perception and memory in healthy volunteers following selective serotonin and norepinephrine reuptake inhibition. *The American Journal of Psychiatry* 161: 1256–1263, 2004.

Hayashi K, Nakao K, Nakamura K. Appetitive and aversive information coding in the primate dorsal raphé nucleus. *Journal of Neuroscience* 35: 6195–6208, 2015.

Herzfeld DJ, Vaswani PA, Marko MK, Shadmehr R. A memory of errors in sensorimotor learning. *Science* 345: 1349–1353, 2014.

Heym J, Trulson ME, Jacobs BL. Raphe unit activity in freely moving cats: effects of phasic auditory and visual stimuli. *Brain Research* 232: 29–39, 1982.

Hornung JP. The human raphe nuclei and the serotonergic system. *Journal of Chemical Neuroanatomy* 26: 331–343, 2003.

Huang KW, Ochandarena NE, Philson AC, Hyun M, Birnbaum JE, Cicconet M, Sabatini BL. Molecular and anatomical organization of the dorsal raphe nucleus. *eLife* 8:e46464, 2019.

Iigaya K, Fonseca MS, Murakami M, Mainen ZF, Dayan P. An effect of serotonergic stimulation on learning rates for rewards apparent after long intertrial intervals. *Nature Communications* 9: 2477, 2018.

Ishimura K, Takeuchi Y, Fujiwara K, Tominaga M, Yoshioka H, Sawada T. Quantitative analysis of the distribution of serotonin-immunoreactive cell bodies in the mouse brain. *Neuroscience Letters* 91: 265–270, 1988.

Ishiwata T, Hasegawa H, Greenwood BN. Involvement of serotonin in the ventral tegmental area in thermoregulation of freely moving rats. *Neuroscience Letters* 653: 71–77, 2017.

Jacobs B, Azmitia E. Structure and function of the brain serotonin system. *Physiological Reviews* 72: 165–229, 1992.

Jacobs BL, Fornal CA. Activity of brain serotonergic neurons in the behaving animal. *Pharmacological Reviews* 43: 563–578, 1991.

Jacobs BL, Fornal CA. Serotonin and behaviour: a general hypothesis. In *Psychopharmacology: the fourth generation of progress*. Raven Press, 1995.

Jensen P, Farago AF, Awatramani RB, Scott MM, Deneris ES, Dymecki SM. Redefining the serotonergic system by genetic lineage. *Nature Neuroscience* 11: 417–419, 2008.

Jones LA, Sun EW, Martin AM, Keating DJ. The ever-changing roles of serotonin. *International journal of Biochemistry and Cell Biology* 125: 105776, 2020.

Kakade S, Dayan P. Acquisition and extinction in autoshaping. *Psychological Review* 109: 533–544, 2002.

Kanen JW, Apergis-Schoute AM, Yellowlees R, Arntz FE, van der Flier FE, Price A, Cardinal RN, Christmas DM, Clark L, Sahakian BJ, Crockett MJ, Robbins TW. Serotonin depletion impairs both Pavlovian and instrumental reversal learning in healthy humans. *bioRxiv* , 2020.

Kao CH, Lee S, Gold JI, Kable JW. Neural encoding of task-dependent errors during adaptive learning. *eLife* 9:e58809, 2020.

Keating DJ, Spencer NJ. What is the role of endogenous gut serotonin in the control of gastrointestinal motility? *Pharmacological Research* 140: 50–55, 2019.

Kennerley SW, Walton ME, Behrens TEJ, Buckley MJ, Rushworth MFS. Optimal decision making and the anterior cingulate cortex. *Nature Neuroscience* 9: 940–947, 2006.

Kiser D, Steemers B, Branchi I, Homberg JR. The reciprocal interaction between serotonin and social behaviour. *Neuroscience and Biobehavioral Reviews* 36: 786–798, 2012.

Kjaerby C, Athilingam J, Robinson SE, Iafrati J, Sohal VS. Serotonin 1B Receptors Regulate Prefrontal Function by Gating Callosal and Hippocampal Inputs. *Cell Reports* 17: 2882–2890, 2016.

Kocsis B, Varga V, Dahan L, Sik A. Serotonergic neuron diversity: identification of raphe neurons with discharges time-locked to the hippocampal theta rhythm. *Proceedings of the National Academy of Sciences of the United States of America* 103: 1059–1064, 2006.

Krugel LK, Biele G, Mohr PN, Li SC, Heekeren HR. Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. *Proceedings of the National Academy of Sciences of the United States of America* 106: 17951–17956, 2009.

Lau B, Glimcher PW. Dynamic response-by-response models of matching behavior in rhesus monkeys. *Journal of Experimental Analysis of Behavior* 84: 555–579, 2005.

Lau B, Glimcher PW. Value representations in the primate striatum during matching behavior. *Neuron* 58: 451–463, 2008.

Lee MD, Clifton PG. Chapter 27 - Role of the serotonergic system in appetite and ingestion control. In *Handbook of the Behavioral Neurobiology of Serotonin*, pp. 469–487. 2020.

Lee SW, Shimojo S, O'Doherty JP. Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 81: 687–699, 2014.

Lee YA, Goto Y. The Roles of Serotonin in Decision-making under Social Group Conditions. *Scientific Reports* 8: 1–11, 2018.

Lefebvre G, Lebreton M, Meyniel F, Bourgeois-Gironde S, Palminteri S. Behavioural and neural characterization of optimistic reinforcement learning. *Nature Human Behaviour* 1: 1–9, 2017.

Lesch KP, Waider J. Serotonin in the modulation of neural plasticity and networks: implications for neurodevelopmental disorders. *Neuron* 76: 175–191, 2012.

Li Y, Dalphin N, Hyland BI. Association with reward negatively modulates short latency phasic conditioned responses of dorsal raphe nucleus neurons in freely moving rats. *Journal of Neuroscience* 33: 5065–5078, 2013.

Li Y, Zhong W, Wang D, Feng Q, Liu Z, Zhou J, Jia C, Hu F, Zeng J, Guo Q, Fu L, Luo M. Serotonin neurons in the dorsal raphe nucleus encode reward signals. *Nature Communications* 7: 10503, 2016.

Lima SQ, Hromádka T, Znamenskiy P, Zador AM. PINP: a new method of tagging neuronal populations for identification during in vivo electrophysiological recording. *PLoS One* 4: e6099, 2009.

Linley SB, Hoover WB, Vertes RP. Pattern of distribution of serotonergic fibers to the orbitomedial and insular cortex in the rat. *Journal of Chemical Neuroanatomy* 48-49: 29–45, 2013.

Linley SB, Olucha-Bordonau F, Vertes RP. Pattern of distribution of serotonergic fibers to the amygdala and extended amygdala in the rat. *The Journal of Comparative Neurology* 525: 116–139, 2017.

Liu Z, Zhou J, Li Y, Hu F, Lu Y, Ma M, Feng Q, Zhang JE, Wang D, Zeng J, Bao J, Kim JY, Chen ZF, El Mestikawy S, Luo M. Dorsal raphe neurons signal reward through 5-HT and glutamate. *Neuron* 81: 1360–1374, 2014.

Lottem E, Banerjee D, Vertechi P, Sarra D, Lohuis MO, Mainen ZF. Activation of serotonin neurons promotes active persistence in a probabilistic foraging task. *Nature Communications* 9: 1000, 2018.

Lottem E, Lörincz ML, Mainen ZF. Optogenetic Activation of Dorsal Raphe Serotonin Neurons Rapidly Inhibits Spontaneous But Not Odor-Evoked Activity in Olfactory Cortex. *Journal of Neuroscience* 36: 7–18, 2016.

Luce RD. *Response Times: Their Role in Inferring Elementary Mental Organization*. Oxford University Press, 1986.

Mammadova-Bach E, Mauler M, Braun A, Duerschmied D. Autocrine and paracrine regulatory functions of platelet serotonin. *Platelets* 29: 541–548, 2018.

Mann JJ, McBride PA, Anderson GM, Mieczkowski TA. Platelet and whole blood serotonin content in depressed inpatients: correlations with acute and life-time psychopathology. *Biological Psychiatry* 32: 243–257, 1992.

Marder E. Neuromodulation of neuronal circuits: back to the future. *Neuron* 76: 1–11, 2012.

Martin AM, Young RL, Leong L, Rogers GB, Spencer NJ, Jessup CF, Keating DJ. The Diverse Metabolic Roles of Peripheral Serotonin. *Endocrinology* 158: 1049–1063, 2017.

Massi B, Donahue CH, Lee D. Volatility Facilitates Value Updating in the Prefrontal Cortex. *Neuron* 99: 598–608.e4, 2018.

Matias S, Lottem E, Dugué GP, Mainen ZF. Activity patterns of serotonin neurons underlying cognitive flexibility. *eLife* 6, 2017.

McCormick DA, Wang Z. Serotonin and noradrenaline excite GABAergic neurones of the guinea-pig and cat nucleus reticularis thalami. *The Journal of Physiology* 442: 235–255, 1991.

McGinty DJ, Harper RM. Dorsal raphe neurons: depression of firing during sleep in cats. *Brain Research* 101: 569–575, 1976.

McGuire JT, Nassar MR, Gold JI, Kable JW. Functionally dissociable influences on learning rate in a dynamic environment. *Neuron* 84: 870–881, 2014.

Michely J, Eldar E, Martin IM, Dolan RJ. A mechanistic account of serotonin's impact on mood. *Nature Communications* 11: 2335, 2020.

Miller KJ, Botvinick MM, Brody CD. Dorsal hippocampus contributes to model-based planning. *Nature Neuroscience* 20: 1269–1276, 2017.

Miller KJ, Ludvig EA, Pezzulo G, Shenhav A. *Goal-Directed Decision Making*, chap. Chapter 18 - Realigning Models of Habitual and Goal-Directed Decision-Making, pp. 407–428. 2018.

Miyazaki K, Miyazaki KW, Doya K. Activation of dorsal raphe serotonin neurons underlies waiting for delayed rewards. *Journal of Neuroscience* 31: 469–479, 2011.

Miyazaki K, Miyazaki KW, Sivori G, Yamanaka A, Tanaka KF, Doya K. Serotonergic projections to the orbitofrontal and medial prefrontal cortices differentially modulate waiting for future rewards. *Science Advances* 6, 2020.

Miyazaki KW, Miyazaki K, Tanaka KF, Yamanaka A, Takahashi A, Tabuchi S, Doya K. Optogenetic activation of dorsal raphe serotonin neurons enhances patience for future rewards. *Current Biology* 24: 2033–2040, 2014.

Monosov IE. How Outcome Uncertainty Mediates Attention, Learning, and Decision-Making. *Trends in Neuroscience* 43: 795–809, 2020.

Montalbano A, Corradetti R, Mlinar B. Pharmacological Characterization of 5-HT1A Autoreceptor-Coupled GIRK Channels in Rat Dorsal Raphe 5-HT Neurons. *PloS one* 10: e0140369, 2015.

Monti JM. Serotonin control of sleep-wake behavior. *Sleep Medicine Reviews* 15: 269–281, 2011.

Morales M, Bloom FE. The 5-HT3 receptor is present in different subpopulations of GABAergic neurons in the rat telencephalon. *Journal of Neuroscience* 17: 3157–3167, 1997.

Morrison SF, Nakamura K. Central neural pathways for thermoregulation. *Frontiers in Bioscience* 16: 74–104, 2011.

Muzerelle A, Scotto-Lomassese S, Bernard JF, Soiza-Reilly M, Gaspar P. Conditional anterograde tracing reveals distinct targeting of individual serotonin cell groups (B5-B9) to the forebrain and brainstem. *Brain Structure and Function* 221: 535–561, 2014.

Nagel G, Szellas T, Huhn W, Kateriya S, Adeishvili N, Berthold P, Ollig D, Hegemann P, Bamberg E. Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proceedings of the National Academy of Sciences of the United States of America* 100: 13940–13945, 2003.

Nakamura K, Matsumoto M, Hikosaka O. Reward-dependent modulation of neuronal activity in the primate dorsal raphe nucleus. *Journal of Neuroscience* 28: 5331–5343, 2008.

Nardou R, Lewis EM, Rothhaas R, Xu R, Yang A, Boyden E, Dölen G. Oxytocin-dependent reopening of a social reward learning critical period with MDMA. *Nature* 569: 116–120, 2019.

Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasly B, Gold JI. Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience* 15: 1040–1046, 2012.

Neugebauer V. Chapter 17 - Serotonin—pain modulation. In *Handbook of the Behavioral Neurobiology of Serotonin*, pp. 309–320. 2020.

Niederkofler V, Asher TE, Okaty BW, Rood BD, Narayan A, Hwa LS, Beck SG, Miczek KA, Dymecki SM. Identification of Serotonergic Neuronal Modules that Affect Aggressive Behavior. *Cell Reports* 17: 1934–1949, 2016.

Ogawa SK, Cohen JY, Hwang D, Uchida N, Watabe-Uchida M. Organization of monosynaptic inputs to the serotonin and dopamine neuromodulatory systems. *Cell Reports* 8: 1105–1118, 2014.

Okaty BW, Commons KG, Dymecki SM. Embracing diversity in the 5-HT neuronal system. *Nature Reviews Neuroscience* 20: 397–424, 2019.

Okaty BW, Freret ME, Rood BD, Brust RD, Hennessy ML, deBairos D, Kim JC, Cook MN, Dymecki SM. Multi-Scale Molecular Deconstruction of the Serotonin Neuron System. *Neuron* 88: 774–791, 2015.

Okaty BW, Sturrock N, Lozoya YE, Chang Y, Senft RA, Lyon KA, Alekseyenko OV, Dymecki SM. A single-cell transcriptomic and anatomic atlas of mouse dorsal raphe Pet1 neurons. *Elife* 9: e55523, 2020.

O'Reilly JX. Making predictions in a changing world—inference, uncertainty, and learning. *Frontiers in Neuroscience* 7: 105, 2013.

Otake K, Kin K, Nakamura Y. Fos expression in afferents to the rat midline thalamus following immobilization stress. *Neuroscience Research* 43: 269–282, 2002.

Palazzo E, de Novellis V, Petrosino S, Marabese I, Vita D, Giordano C, Di Marzo V, Mangoni GS, Rossi F, Maione S. Neuropathic pain and the endocannabinoid system in the dorsal raphe: pharmacological treatment and interactions with the serotonergic system. *European Journal of Neuroscience* 24: 2011–2020, 2006.

Palazzo E, Genovese R, Mariani L, Siniscalco D, Marabese I, de Novellis V, Rossi F, Maione S. Metabotropic glutamate receptor 5 and dorsal raphe serotonin release in inflammatory pain in rat. *European Journal of Pharmacology* 492: 169–176, 2004.

Parent M, Descarries L. Chapter 4 - Ultrastructure of the serotonin innervation in mammalian central nervous system. In *Handbook of the Behavioral Neurobiology of Serotonin*, pp. 49–90. 2020.

Park SB, Coull JT, McShane RH, Young AH, Sahakian BJ, Robbins TW, Cowen PJ. Tryptophan depletion in normal volunteers produces selective impairments in learning and memory. *Neuropharmacology* 33: 575–588, 1994.

Payzan-LeNestour E, Bossaerts P. Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS computational biology* 7, 2011.

Payzan-LeNestour E, Dunne S, Bossaerts P, O'Doherty JP. The neural representation of unexpected uncertainty during value-based decision making. *Neuron* 79: 191–201, 2013.

Pearce JM, Hall G. A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review* 87: 532–552, 1980.

Penington NJ, Kelly JS, Fox AP. Whole-cell recordings of inwardly rectifying K+ currents activated by 5-HT1A receptors on dorsal raphe neurones of the adult rat. *The Journal of Physiology* 469: 387–405, 1993.

Pollak Dorocic I, Fürth D, Xuan Y, Johansson Y, Pozzi L, Silberberg G, Carlén M, Meletis K. A whole-brain atlas of inputs to serotonergic neurons of the dorsal and median raphe nuclei. *Neuron* 83: 663–678, 2014.

Preuschoff K, Bossaerts P. Adding prediction risk to the theory of reward learning. *Annals of the New York Academy of Sciences* 1104: 135–146, 2007.

Preuschoff K, Bossaerts P, Quartz SR. Neural differentiation of expected reward and risk in human subcortical structures. *Neuron* 51: 381–390, 2006.

Preuschoff K, Quartz SR, Bossaerts P. Human insula activation reflects risk prediction errors as well as risk. *Journal of Neuroscience* 28: 2745–2752, 2008.

Prinz AA, Bucher D, Marder E. Similar network activity from disparate circuit parameters. *Nature Neuroscience* 7: 1345–1352, 2004.

Prouty EW, Chandler DJ, Waterhouse BD. Neurochemical differences between target-specific populations of rat dorsal raphe projection neurons. *Brain Research* 1675: 28–40, 2017.

Puig MV, Artigas F, Celada P. Modulation of the activity of pyramidal neurons in rat prefrontal cortex by raphe stimulation in vivo: involvement of serotonin and GABA. *Cerebral Cortex* 15: 1–14, 2005.

Quiroga RQ, Nadasdy Z, Ben-Shaul Y. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Computation* 16: 1661–1687, 2004.

Ranade SP, Mainen ZF. Transient firing of dorsal raphe neurons encodes diverse and specific sensory, motor, and reward events. *Journal of Neurophysiology* 102: 3026–3037, 2009.

Rapport M, Green A, Page I. Serum vasoconstrictor, serotonin; isolation and characterization. *The Journal of Biological Chemistry* 176: 1243–1251, 1948.

Rasmussen K, Strecker RE, Jacobs BL. Single unit response of noradrenergic, serotonergic and dopaminergic neurons in freely moving cats to simple sensory stimuli. *Brain Research* 369: 336–340, 1986.

Ray RS, Corcoran AE, Brust RD, Kim JC, Richerson GB, Nattie E, Dymecki SM. Impaired respiratory and body temperature control upon acute serotonergic neuron inhibition. *Science* 333: 637–642, 2011.

Ren J, Friedmann D, Xiong J, Liu CD, Ferguson BR, Weerakkody T, DeLoach KE, Ran C, Pun A, Sun Y, Weissbourd B, Neve RL, Huguenard J, Horowitz MA, Luo L. Anatomically defined and functionally distinct dorsal raphe serotonin sub-systems. *Cell* 175: 472–487.e20, 2018.

Ren J, Isakova A, Friedmann D, Zeng J, Grutzner SM, Pun A, Zhao GQ, Kolluru SS, Wang R, Lin R, et al. Single-cell transcriptomes and whole-brain projections of serotonin neurons in the mouse dorsal and median raphe nuclei. *Elife* 8: e49424, 2019.

Ribeiro-do Valle L, Fornal C, Litto W, Jacobs B. Serotonergic dorsal raphe unit activity related to feeding/grooming behaviors in cats. In *Society for Neuroscience Abstracts*, vol. 15, p. 1. 1989.

Roberts C, Sahakian BJ, Robbins TW. Psychological mechanisms and functions of 5-HT and SSRIs in potential therapeutic change: Lessons from the serotonergic modulation of action selection, learning, affect, and social cognition. *Neuroscience and Biobehavioral Reviews* 119: 138–167, 2020.

Rogers RD, Blackshaw AJ, Middleton HC, Matthews K, Hawtin K, Crowley C, Hopwood A, Wallace C, Deakin JF, Sahakian BJ, Robbins TW. Tryptophan depletion impairs stimulus-reward learning while methylphenidate disrupts attentional control in healthy young adults: implications for the monoaminergic basis of impulsive behaviour. *Psychopharmacology* 146: 482–491, 1999.

Rutledge RB, Skandali N, Dayan P, Dolan RJ. A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences of the United States of America* 111: 12252–12257, 2014.

Samejima K, Ueda Y, Doya K, Kimura M. Representation of action-specific reward values in the striatum. *Science* 310: 1337–1340, 2005.

Santana N, Artigas F. Expression of Serotonin2C Receptors in Pyramidal and GABAergic Neurons of Rat Prefrontal Cortex: A Comparison with Striatum. *Cerebral Cortex* 27: 3125–3139, 2017.

Saxena PR, Tangri KK, Bhargava KP. Identification of acetylcholine, histamine, and 5-hydroxytryptamine in Girardinia heterophylla (Decne.). *Canadian Journal of Physiology and Pharmacology* 44: 621–627, 1966.

Schmitzer-Torbert N, Jackson J, Henze D, Harris K, Redish A. Quantitative measures of cluster quality for use in extracellular recordings. *Neuroscience* 131: 1–11, 2005.

Schultz W, Dayan P, Montague P. A neural substrate of prediction and reward. *Science* 275: 1593–1599, 1997.

Schultz W, Stauffer WR, Lak A. The phasic dopamine signal maturing: from reward via behavioural activation to formal economic utility. *Current Opinion in Neurobiology* 43: 139–148, 2017.

Schweimer J, Mallet N, Sharp T, Ungless M. Spike-timing relationship of neurochemically-identified dorsal raphe neurons during cortical slow oscillations. *Neuroscience* 196: 115–123, 2011.

Schweimer J, Ungless M. Phasic responses in dorsal raphe serotonin neurons to noxious stimuli. *Neuroscience* 171: 1209–1215, 2010.

Segarra N, Metastasio A, Ziauddeen H, Spencer J, Reinders NR, Dudas RB, Arrondo G, Robbins TW, Clark L, Fletcher PC, Murray GK. Abnormal Frontostriatal Activity During Unexpected Reward Receipt in Depression and Schizophrenia: Relationship to Anhedonia. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology* 41: 2001–2010, 2016.

Seillier L, Lorenz C, Kawaguchi K, Ott T, Nieder A, Pourriahi P, Nienborg H. Serotonin decreases the gain of Vvsual responses in awake macaque V1. *Journal of Neuroscience* 37: 11390–11405, 2017.

Sengupta A, Holmes A. A Discrete Dorsal Raphe to Basal Amygdala 5-HT Circuit Calibrates Aversive Memory. *Neuron* 103: 489–505.e7, 2019.

Seo C, Guru A, Jin M, Ito B, Sleezer BJ, Ho YY, Wang E, Boada C, Krupa NA, Kullakanda DS, Shen CX, Warden MR. Intense threat switches dorsal raphe serotonin neurons to a paradoxical operational mode. *Science* 363: 538–542, 2019.

Seo H, Lee D. Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *Journal of Neuroscience* 27: 8366–8377, 2007.

Sheldon PW, Aghajanian GK. Excitatory responses to serotonin (5-HT) in neurons of the rat piriform cortex: evidence for mediation by 5-HT1C receptors in pyramidal cells and 5-HT2 receptors in interneurons. *Synapse* 9: 208–218, 1991.

Shima K, Tanji J. Role for cingulate motor area cells in voluntary movement selection based on reward. *Science* 282: 1335–1338, 1998.

Shimegi S, Kimura A, Sato A, Aoyama C, Mizuyama R, Tsunoda K, Ueda F, Araki S, Goya R, Sato H. Cholinergic and serotonergic modulation of visual information processing in monkey V1. *Journal of Physiology* 110: 44–51, 2016.

Simansky KJ. Serotonergic control of the organization of feeding and satiety. *Behavioural Brain Research* 73: 37–42, 1996.

Smith TA. The occurrence, metabolism and functions of amines in plants. *Biological Reviews of the Cambridge Philosophical Society* 46: 201–241, 1971.

Soltani A, Izquierdo A. Adaptive learning under expected and unexpected uncertainty. *Nature Reviews Neuroscience* 20: 635–644, 2019.

Soubrié P. Reconciling the role of central serotonin neurons in human and animal behavior. *Behavioral and Brain Sciences* 9: 319–335, 1986.

Sparks DW, Tian MK, Sargin D, Venkatesan S, Intson K, Lambe EK. Opposing Cholinergic and Serotonergic Modulation of Layer 6 in Prefrontal Cortex. *Frontiers in Neural Circuits* 11: 107, 2017.

Spencer NJ, Sia TC, Brookes SJ, Costa M, Keating DJ. CrossTalk opposing view: 5-HT is not necessary for peristalsis. *The Journal of Physiology* 593: 3229–3231, 2015.

Steinbusch HW. Distribution of serotonin-immunoreactivity in the central nervous system of the rat-cell bodies and terminals. *Neuroscience* 6: 557–618, 1981.

Stephens EK, Avesar D, Gulledge AT. Activity-dependent serotonergic excitation of callosal projection neurons in the mouse prefrontal cortex. *Frontiers in Neural Circuits* 8: 97, 2014.

Stephens EK, Baker AL, Gulledge AT. Mechanisms Underlying Serotonergic Excitation of Callosal Projection Neurons in the Mouse Medial Prefrontal Cortex. *Frontiers in Neural Circuits* 12: 2, 2018.

Sugrue LP, Corrado GS, Newsome WT. Matching behavior and the representation of value in the parietal cortex. *Science* 304: 1782–1787, 2004.

Sutton RS, Barto AG. *Reinforcement Learning: An Introduction*. MIT Press Cambridge, 1998.

Svensson E, Apergis-Schoute J, Burnstock G, Nusbaum MP, Parker D, Schiöth HB. General Principles of Neuronal Co-transmission: Insights From Multiple Model Systems. *Frontiers in Neural Circuits* 12: 117, 2018.

Szabo ST, Blier P. Functional and pharmacological characterization of the modulatory role of serotonin on the firing activity of locus coeruleus norepinephrine neurons. *Brain Research* 922: 9–20, 2001.

Tecott LH, Maricq AV, Julius D. Nervous system distribution of the serotonin 5-HT3 receptor mRNA. *Proceedings of the National Academy of Sciences of the United States of America* 90: 1430–1434, 1993.

Teissier A, Chemiakine A, Inbar B, Bagchi S, Ray RS, Palmiter RD, Dymecki SM, Moore H, Ansorge MS. Activity of Raphé Serotonergic Neurons Controls Emotional Behaviors. *Cell Reports* 13: 1965–1976, 2015.

Thompson KJ, Khajehali E, Bradley SJ, Navarrete JS, Huang XP, Slocum S, Jin J, Liu J, Xiong Y, Olsen RHJ, Diberto JF, Boyt KM, Pina MM, Pati D, Molloy C, Bundgaard C, Sexton PM, Kash TL, Krashes MJ, Christopoulos A, Roth BL, Tobin AB. DREADD Agonist 21 Is an Effective Agonist for Muscarinic-Based DREADDs. *ACS Pharmacology and Translational Science* 1: 61–72, 2018.

Tobler PN, Fiorillo CD, Schultz W. Adaptive coding of reward value by dopamine neurons. *Science* 307: 1642–1645, 2005.

Trulson ME, Jacobs BL. Raphe unit activity in freely moving cats: correlation with level of behavioral arousal. *Brain Research* 163: 135–150, 1979.

Tsutsui KI, Grabenhorst F, Kobayashi S, Schultz W. A dynamic code for economic object valuation in prefrontal cortex neurons. *Nature Communications* 7: 12554, 2016.

Törk I. Anatomy of the serotonergic system. *Annals of the New York Academy of Sciences* 600: 9–34, 1990.

Törk I, Hornung JP. Raphe Nuclei and the Serotonergic System. In *The Human Nervous System*. Academic Press, 1990.

Vandermaelen CP, Aghajanian GK. Electrophysiological and pharmacological characterization of serotonergic dorsal raphe neurons recorded extracellularly and intracellularly in rat brain slices. *Brain Research* 289: 109–119, 1983.

Vanhoutte PM. Serotonin and the vascular wall. *International Journal of Cardiology* 14: 189–203, 1987.

Vertes RP. A PHA-L analysis of ascending projections of the dorsal raphe nucleus in the rat. *Journal of Comparative Neurology* 313: 643–668, 1991.

Vertes RP, Fortin WJ, Crane AM. Projections of the median raphe nucleus in the rat. *The Journal of Comparative Neurology* 407: 555–582, 1999.

Vertes RP, Linley SB. Comparison of projections of the dorsal and median raphe nuclei, with some functional considerations. *International Congress Series* 1304: 98 – 120, 2007.

Vertes RP, Linley SB. *Efferent and afferent connections of the dorsal and median raphe nuclei in the rat*, pp. 69–102. Basel: Birkhäuser Basel, 2008.

Vertes RP, Linley SB, Hoover WB. Pattern of distribution of serotonergic fibers to the thalamus of the rat. *Brain Structure and Function* 215: 1–28, 2010.

Vialli M, Erspamer V. Ricerche sul secreto delle cellule enterocromaffini. *Journal of Cell Research and Microscopic Anatomy* 27: 81–99, 1937.

Vilaró MT, Cortés R, Mengod G, Hoyer D. Chapter 6 - Distribution of 5-HT receptors in the central nervous system: an update. In *Handbook of the Behavioral Neurobiology of Serotonin*, pp. 121–146. 2020.

Vrieze E, Pizzagalli DA, Demyttenaere K, Hompes T, Sienaert P, de Boer P, Schmidt M, Claes S. Reduced reward learning predicts outcome in major depressive disorder. *Biological Psychiatry* 73: 639–645, 2013.

Walther DJ, Peter JU, Winter S, Höltje M, Paulmann N, Grohmann M, Vowinckel J, Alamo-Bethencourt V, Wilhelm CS, Ahnert-Hilger G, Bader M. Serotonylation of small GTPases is a signal transduction pathway that triggers platelet alpha-granule release. *Cell* 115: 851–862, 2003.

Wang AY, Miura K, Uchida N. The dorsomedial striatum encodes net expected return, critical for energizing performance vigor. *Nature Neuroscience* 16: 639–647, 2013.

Wang JX, Kurth-Nelson Z, Kumaran D, Tirumala D, Soyer H, Leibo JZ, Hassabis D, Botvinick M. Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience* 21: 860–868, 2018.

Weissbourd B, Ren J, DeLoach KE, Guenthner CJ, Miyamichi K, Luo L. Presynaptic partners of dorsal raphe serotonergic and GABAergic neurons. *Neuron* 83: 645–662, 2014.

Wier J, Tyler VE. Quantitative determination of serotonin in Panaeolus species. *Journal of Pharmaceutical Sciences* 52: 419–422, 1963.

Winstanley CA, Theobald DEH, Dalley JW, Glennon JC, Robbins TW. 5-HT2A and 5-HT2C receptor antagonists have opposing effects on a measure of impulsivity: interactions with global 5-HT depletion. *Psychopharmacology* 176: 376–385, 2004.

Wise CD, Berger BD, Stein L. Serotonin: a possible mediator of behavioral suppression induced by anxiety. *Diseases of the Nervous System* 31: 34–37, 1970.

Wise CD, Berger BD, Stein L. Evidence of -noradrenergic reward receptors and serotonergic punishment receptors in the rat brain. *Biological Psychiatry* 6: 3–21, 1973.

Wittmann MK, Fouragnan E, Folloni D, Klein-Flügge MC, Chau BK, Khamassi M, Rushworth MF. Global reward state affects learning and activity in raphe nucleus and anterior insula in monkeys. *Nature Communications* 11: 1–17, 2020.

Wood RM, Rilling JK, Sanfey AG, Bhagwagar Z, Rogers RD. Effects of Tryptophan Depletion on the Performance of an Iterated Prisoner's Dilemma Game in Healthy Adults. *Neuropsychopharmacology* 31: 1075–1084, 2006.

Xie C, Jia T, Rolls ET, Robbins TW, Sahakian BJ, Zhang J, Liu Z, Cheng W, Luo Q, Zac Lo CY, Wang H, Banaschewski T, Barker GJ, Bokde ALW, Büchel C, Quinlan EB, Desrivières S, Flor H, Grigis A, Garavan H, Gowland P, Heinz A, Hohmann S, Ittermann B, Martinot JL, Paillère Martinot ML, Nees F, Orfanos DP, Paus T, Poustka L, Fröhner JH, Smolka MN, Walter H, Whelan R, Schumann G, Feng J, Consortium IMAGEN. Reward Versus Nonreward Sensitivity of the Medial Versus Lateral Orbitofrontal Cortex Relates to the Severity of Depressive Symptoms. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* , 2020.

Young RL, Lumsden AL, Keating DJ. Gut Serotonin Is a Regulator of Obesity and Metabolism. *Gastroenterology* 149: 253–255, 2015.

Yu AJ, Dayan P. Uncertainty, neuromodulation, and attention. *Neuron* 46: 681–692, 2005.

Yuan Q, Lin F, Zheng X, Sehgal A. Serotonin modulates circadian entrainment in Drosophila. *Neuron* 47: 115–127, 2005.

Zhang K, Chen CD, Monosov IE. Novelty, Salience, and Surprise Timing Are Signaled by Neurons in the Basal Forebrain. *Current Biology* 29: 134–142.e3, 2019.

Zhong W, Li Y, Feng Q, Luo M. Learning and stress shape the reward response patterns of serotonin neurons. *Journal of Neuroscience* 37: 8863–8875, 2017.

Zhou J, Jia C, Feng Q, Bao J, Luo M. Prospective coding of dorsal raphe reward signals by the orbitofrontal cortex. *Journal of Neuroscience* 35: 2717–2730, 2015.

Zhou L, Liu MZ, Li Q, Deng J, Mu D, Sun YG. Organization of Functional Long-Range Circuits Controlling the Activity of Serotonergic Neurons in the Dorsal Raphe Nucleus. *Cell Reports* 20: 1991–1993, 2017.

Zhuang X, Masson J, Gingrich JA, Rayport S, Hen R. Targeted gene expression in dopamine and serotonin neurons of the mouse brain. *Journal of Neuroscience Methods* 143: 27–32, 2005.

Appendix I

Hierarchical Bayesian approach to model fitting

Behavior and neuron firing models in this manuscript were fit to the data using a hierarchical Bayesian modeling approach enabled by the probabilistic programming language Stan (https://mc-stan.org/) along with the MATLAB interface, MatlabStan (https://mc-stan.org/users/interfaces/matlab-stan). This approach performs full Bayesian statistical inference on model parameters which will be described in this appendix. Fitting in this way has several advantages over the maximum likelihood estimation (MLE) approach to model fitting.

MLE approach to model fitting

The goal of MLE is to search parameter space to find parameter value estimates that maximize the likelihood function:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta; x),$$

where θ is the set of parameters, Θ is the range of possible values for each parameter, and $\hat{\theta}$ is the estimated set of parameters that make the data, x, most probable by maximizing the likelihood function, $\mathcal{L}$. For many models, such as reinforcement learning algorithms that we use here, the likelihood function is nonlinear because changes in parameters are not proportional to the resulting change in likelihood. As such, the above equation cannot be

solved explicitly—there is no closed form solution. Instead, solver-based approaches are used to systematically explore parameter space and optimize the likelihood function.

The output of this method is point (single value) estimates of the parameters. But in trying to find the maximum likelihood, MLE samplers can get stuck in local maxima while exploring parameter space, yielding parameter values that are not the desired estimates. MLE can also be biased when sample sizes are small and this can be exacerbated by the nature of the data as well as the formulation of the model that is being fit to those data. In the dynamic foraging behavior, for example, we model choice as a function of the relative values of actions. The model describes the latent process by which these values are learned in order to drive the observed behavior. As a consequence of the design of the task, mice switch their choices infrequently, which makes estimation of the learning rates difficult. Suppose a mouse never switches their choice for an entire session. Even if the mouse completes many trials, how can we quantify the amount that the mouse is learning from no reward outcomes if their behavior is not modified by them? Another example of non-identifiability, made worse by insufficient data, occurs when changes in distinct parameters in a model have similar effects on the likelihood. For example, the inverse temperature parameter, which controls the greediness of the decision function, has been shown to be inversely correlated with the learning rate parameter (Seo and Lee, 2007; Gershman, 2016). One potential workaround is to concatenate data, but this averages out potentially meaningful variability across sessions. And this variability could be relevant to comparisons between model variables and observations of neural activity in a single session. So, fitting to smaller numbers of samples (e.g., individual sessions) can cause overfitting to noisy data, while fitting to pooled data (e.g., across sessions) can cause underfitting to meaningful variability in behavior.

Hierarchical Bayesian approach to model fitting

The hierarchical Bayesian approach deals with this problem by allowing for principled assumptions to influence fitting in a statistically sound way. The Bayesian approach generally

computes the posterior probability distributions for parameters according to Bayes rule:

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)},$$

which can be read as

$$Posterior = \frac{Likelihood * Prior}{Normalization}.$$

In this way, Bayesian inference considers two sources of information: the information from your data contained in the likelihood and the information in the prior. The prior may be uninformative, contain information according to some principle, or may be derived from past information (e.g., previous experiments). The likelihood and the prior are probability *distributions* which can be described as allocating credibility across possible values. The posterior is determined exactly from both sources of information, so both are important to consider. To generate meaningful posteriors, there must be enough data and those data must be informative relative to the parameters of the proposed model. As long as the priors are not ill-defined (e.g., zero probability density for relevant values), with infinite data the posterior will asymptotically converge to a Gaussian distribution that is independent of the prior. For finite data, how strongly informative your prior is will determine its contribution to the posterior.

Putting priors on parameters allows for the incorporation of information beyond the data to which the model is being fit. Some have advocated for "empirical priors" that synthesize information about model parameters from previous studies (Gershman, 2016). This evidence is used to construct weakly informative priors to influence parameter estimates in a principled fashion. With sufficient data and a reasonable model, uninformative priors that have uniform probability over all possible values can be used. MLE is actually a special case of Bayesian inference in which the priors on the parameters are uninformative in this way and the $\hat{\theta}$ are the maximum *a posteriori* estimates of the posterior probability distributions of parameters. However, the full posteriors on parameters estimated using the Bayesian approach have the advantage of accounting for uncertainty in those estimates.

The Bayesian approach can also address the limitations of MLE in dealing with the under- and overfitting dilemma by leveraging hierarchical model structures that permit partial pooling of data. In these models, it is assumed that information about the data being fit is available at multiple levels. In our case, this strategy allows us to model all the data for a single mouse at once, without sacrificing session-to-session variability. At the level of individual sessions for one mouse, we still modeled choice behavior as a function of the reinforcement learning algorithm. At the level of the mouse, we modeled the assumption that mice tend to behave somewhat consistently across sessions. The session-level parameters then, are essentially modeled as being drawn from mouse-level Gaussian distributions for each parameter. The mouse-level parameters constrain session-level parameters by allowing behavior from individual sessions to inform parameter estimates for the others, while simultaneously allowing for some variability across those sessions.

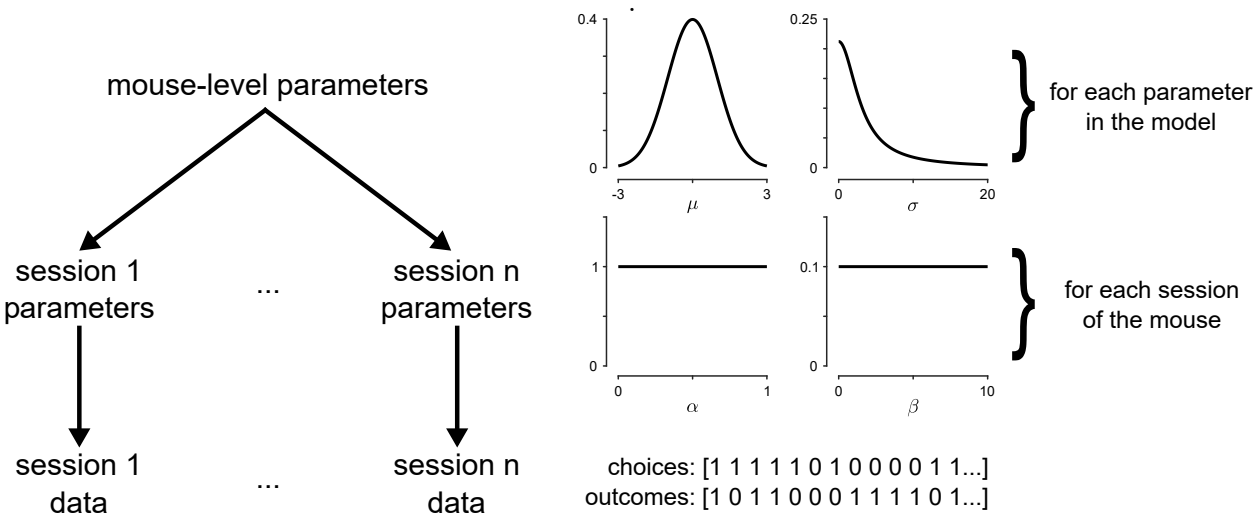

Figure I-1: **Hierarchical Bayesian model.** Schematic of the Hierarchical Bayesian model construction for the simplest Q-learning algorithm. Mouse-level parameters govern session-level parameters to mitigate over-fitting while accommodating variability. Probability density functions show that the priors on session-level parameters are uniform and thus, uninformative. Probability density functions for the priors on mouse-level parameters (that determine the properties of the mouse-level distributions) are weakly informative. Right choices are coded as 1 and left choices as 0. Outcomes are coded as 1 for a reward and 0 for no reward.

Another advantage of this hierarchical structure is that informative priors on session-level parameters are unnecessary because substantial amounts of behavior can be modeled at once.

Thus, we used priors with uniform probabilities across all possible values for each session-level parameter. In other words, no assumptions about these parameter values are built into the model, allowing the likelihood (and therefore the data) to dominate the posteriors. For the parameters that determine mouse-level distributions, we use weakly informative priors that are chosen to regularize the data without ruling out meaningful variability (Gelman, 2006).

An example of a hierarchical Bayesian model

An example can help to show how these models work and how they are implemented. In this example, the simplest form of the Q-learning algorithm will be used to describe foraging behavior in a task with two actions available to the behaving animal. This reinforcement learning model estimates action values ($Q_l(t)$ and $Q_r(t)$) on each trial in order to generate choices. Choices are described by a random variable, $c(t)$, corresponding to left or right choice, $c(t) \in \{l, r\}$. The value of a choice is updated as a function of the reward prediction error, $\delta(t)$. The rate at which this learning occurs is controlled by the learning rate parameter α. For example, if the left spout was chosen, then

$$\delta(t) = R(t) - Q_l(t)$$
$$Q_l(t+1) = Q_l(t) + \alpha\delta(t).$$

The Q-values are used to generate choice probabilities through a softmax decision function (Daw et al., 2006):

$$P(c(t) = r) = \frac{1}{1 + e^{-\beta(Q_r(t) - Q_l(t))}},$$
$$P(c(t) = l) = 1 - P(c(t) = r),$$

where β, the "inverse temperature" parameter, controls the steepness of the sigmoidal function. In other words, β controls the level of exploration versus exploitation with respect to the relative action values. So in the model there are two parameters that need to be estimated

for each session, α and β, as well as two sets of mean and variance parameters that define the mouse-level distributions for the session-level parameters.

Understanding some of how Stan works helps to explain how priors on parameters are defined. Computing the posterior ($P(\theta|x)$ given $\theta = (\theta_1, ...\theta_d)$) is incredibly difficult in high dimensions (d), when a model has many parameters. Markov chain Monte Carlo methods have been designed to sample intelligently from simulated posteriors to numerically approximate the computation. Stan uses a Hamiltonian Markov chain Monte Carlo with an adaptive No-U-Turn sampler to obtain posterior simulations in order to maximize the objective function: the function whose value we want to maximize.

In our model, we want to predict choices, which are coded as binary values ($c_r = 1$ and $c_l = 0$), as a function of the choice probability determined by the model. Thus, we can define the objective function using a Bernoulli distribution. As such, the probability mass function is

$$P(c|\rho) = \begin{cases} \rho & \text{if } c = 1, \text{ and} \\ 1 - \rho & \text{if } c = 0, \end{cases}$$

where ρ is the probability of a right choice. This is the function by which the posterior (joint distribution of parameters) will be evaluated in the program.

Stan simulates the joint distribution of parameters by defining a log probability function over log-transformed parameters. This transformation is necessary so that the parameters are unconstrained. But with bounded parameters (e.g., $\alpha \in [0, 1]$), boundaries are transformed to be unconstrained using a logit transformation. Consequently, as the parameter moves towards the boundaries, the logit-transformed value moves towards positive or negative infinity. Such boundaries result in increased computational difficulty for the sampler. Fortunately, parameters can be transformed in the model so that the parameters being sampled exist on a more manageable scale while the parameters used in the model are bounded. For our bounded, session-level parameters we adopted this strategy. The priors for these parameters

are defined as normal distributions, $\theta \sim \mathcal{N}(0, 1)$, but are transformed to a bounded, uniform distribution with an approximation of the cumulative density function of the standard normal distribution, $\Phi(\theta)$. For the inverse temperature parameter, β, we can scale this uniform distribution up to reasonable values. Theoretically $\beta \in [0, \infty]$ but empirically we never found any estimates greater than ~ 8 using large upper bounds, so we used 10 as an upper bound.

Weakly informative priors were also placed on the parameters that defined the mouse-level distribution of session-level parameters. The prior for the mean of each parameter distribution is given by $\mu \sim \mathcal{N}(0, 1)$ and the variance by $\sigma \sim \text{half-Cauchy}(0, 3)$.

The data are parsed in MATLAB then fit using the Stan code on the next page.

```
data {                                  //declare data
  int<lower=1> N;                       //number of sessions
  int<lower=1> T;                       //maximum number of trials across sessions
  int<lower=1, upper=T> Tsesh[N];       //array of session lengths
  int<lower=0, upper=2> choice[N, T];   //choice matrix, right = 1, left = 0
  real outcome[N, T];                   //outcome matrix, rewards = 1, no rewards = 0
}
parameters {                            //declare parameters
  //hyper(mouse)-parameters
  vector[2] mu_p;
  vector<lower=0>[2] sigma;

  //session-level raw parameters
  vector[N] a_pr;                       //learning rate
  vector[N] beta_pr;                    //inverse temperature
}
transformed parameters {                //transform session-level raw parameters
  vector<lower=0, upper=1>[N] a;
  vector<lower=0, upper=10>[N] beta;

  for (n in 1:N) {
    a[n]   = Phi_approx(mu_p[1] + sigma[1] * a_pr[n]);
    beta[n] = Phi_approx(mu_p[2] + sigma[2] * beta_pr[n]) * 10;
  }
}
model {                                 //define priors and model
  mu_p    ~ normal(0, 1);               //priors for hyper-parameters
  sigma   ~ cauchy(0, 3);
  a_pr    ~ normal(0, 1);               //priors for individual parameters
  beta_pr ~ normal(0, 1);

  for (n in 1:N) {                      //session loop
    vector[2] Q;                        //expected values, [Q_l, Q_r]
    real RPE;                           //reward prediction error
    vector[Tsesh[n]] Qdiff;             //relative value

    Q = rep_vector(0.0, 2);

    for (t in 1:(Tsesh[n])) {           //trial loop
      Qdiff[t] = Q[2] - Q[1];

      if (choice[n,t] == 1) {           //right choice
        RPE = outcome[n, t] - Q[2];     //calculate reward prediction error
        Q[2] = Q[2] + a[n] * RPE;       //learning rule
      }else{                            //left choice
        PE = outcome[n, t] - Q[1];
        Q[1] = Q[1] + a[n] * RPE;
      }
    }
    //softmax and objective function
    choice[n, 1:Tsesh[n]] ~ bernoulli( 1 / (1 + exp(beta[n] * Qdiff)) );
  }
}
```

Curriculum Vitae **February, 2021**

COOPER D. GROSSMAN

2000 E Lombard St
Baltimore, Maryland 21231 USA
+1.925.699.3136
cgrossm4@jhmi.edu

EDUCATION

2015–Present **Ph.D. in Neuroscience candidate**
Solomon H. Snyder Department of Neuroscience
Johns Hopkins University School of Medicine

2013–2014 **Undergraduate coursework**
University of California, Berkeley Extension Program

2008–2012 **B.A. in Philosophy**
Minors in Natural Science and Medical Anthropology
College of Letters, Arts, and Science
University of Southern California

RESEARCH

August 2016–Present **Graduate Student in the lab of Dr. Jeremiah Cohen**
Solomon H. Snyder Department of Neuroscience
Johns Hopkins University School of Medicine
Along with modeling and manipulation approaches, I am using extracellular electrophysiology techniques paired with optogenetic identification to record from serotonin neurons in mice performing a dynamic foraging task in order to better understand the role of this population in learning and decision making

October 2014–July 2015 **Research Assistant in the lab of Dr. Vikaas Sohal**
Weill Institute for Neurosciences
University of California, San Francisco
Conducted *in vitro* whole cell recordings with optogenetic manipulation and pharmacology to examine the the effects of cholinergic modulation on the recurrent network activity of layer 5 pyramidal neurons in prefrontal cortex, helped various lab members with in vivo calcium imaging, optogenetic manipulation during behavior, patch clamp recording, various behavioral assays, and experiment design, as well as genotyping, mouse colony management, mouse breeding, and assisting the lab manager with managerial tasks.

December 2013–
October 2014

Volunteer Research Assistant in the lab of Dr. Patricia Janak
Weill Institute for Neurosciences
University of California, San Francisco
Worked closely with Dr. E. Zayra Millan to look at the limbic circuitry involved in regulating reward-seeking behavior in the presence and absence of conditioned cues, helped various lab members with optimizing and running genotyping processes, performed different histological tasks, helped to organize and maintain the mouse colony, performed animal drug infusions, and conducted behavior and optogenetic experiments.

TECHNIQUES & SKILLS

Behavior	learning and decision making task design and execution; addiction behavior design and execution; behavioral apparatus design and engineering
Theory & Mathematical Modeling	reinforcement learning models; Bayesian decision models; maximum likelihood estimation for model fitting; hierarchical Bayesian approaches to model fitting
Electrophysiology Techniques	*in vitro* whole cell recordings; *in vivo* extracellular single neuron recordings with tetrodes, electrodes, and silicon probes
Manipulation Techniques	systemic pharmacology; targeted brain region pharmacology; optogenetic activation and inhibition of cell bodies and axon terminals; systemic chemogenetic inhibition; targeted brain region chemogenetic inhibition
Surgical Techniques	targeted virus injections; headplate attachment for head-fixed behavior; optic fiber implantation; drug infusion cannulae implantation; microdrive design and fabrication; microdrive with tetrode implantation
Histological Techniques	tissue perfusions; tissue sectioning; antibody staining; microscopy and image acquisition
Computing & Other Skills	Windows and Macintosh platform software and hardware; MATLAB programming; Stan and cmdstan programming; basic R programming; Arduino and C++ programming; LaTeX programming; 3D design and printing; proper waste disposal practices and safety training; ethical use of vertebrate animals (mice) in research training

TEACHING

January 2017– May 2017	**Teaching Assistant** Johns Hopkins University School of Medicine Neuroscience and Cognition II

COURSES

Ocotober 2018	**Mechanistic Cognitive Neuroscience Workshop** HHMI Janelia Research Campus
Fall 2012	**Graduate Seminar on the Philosophy of Perception** Department of Philosophy University of California, Berkeley

PUBLICATIONS

- **Grossman CD**, Bari BA, Cohen JY. Serotonin neurons modulate learning rate through uncertainty. *BioRxiv*. 2020 Oct 24. 10.24.353508.

- Cohen JY and **Grossman CD**. "Dorsal raphe serotonergic neurons regulate behavior on multiple timescales." Handbook of Behavioral Neuroscience. Vol. 31. Elsevier, 2020. 521-529.

- Lewis EM, Stein-O'Brien GL, Patino AV, Nardou R, **Grossman CD**, Brown M, Bangamwabo B, Ndiaye N, Giovinazzo D, Dardani I, Jiang C, Goff LA, Dölen G. Parallel Social Information Processing Circuits Are Differentially Impacted in Autism. *Neuron*. 2020 Nov 25; 108(4):659-675.e6.

- Bari BA, **Grossman CD**, Lubin EE, Rajagopalan AE, Cressy JI, Cohen JY. Stable Representations of Decision Variables for Flexible Behavior. *Neuron*. 2019 Sep 4; 103(5):922-933.e7.

- Millan EZ, Reese RM, **Grossman CD**, Chaudhri N, Janak PH. Nucleus Accumbens and Posterior Amygdala Mediate Cue-Triggered Alcohol Seeking and Suppress Behavior During the Omission of Alcohol-Predictive Cues. *Neuropsychopharmacology*. 2015 Apr 15; 40 (11): 2555-65.

P R E S E N T A T I O N S

Talks

April 2022	International Society for Serotonin Research (ISSR) Biennial Meeting. *Cancún, Mexico*.
February 2021	Computational and Systems Neuroscience (Cosyne). *Virtual*.
May 2019	Johns Hopkins Neuroscience Department Lab Lunch. *Baltimore, MD*.
February 2019	Computational and Systems Neuroscience (Cosyne). *Lisbon, Portugal*.
October 2018	Janelia Mechanistic Cognitive Neuroscience Workshop. *Ashburn, Virginia*.

Posters

April 2022	International Society for Serotonin Research (ISSR) Biennial Meeting. *Cancún, Mexico*.
February 2020	Computational and Systems Neuroscience (Cosyne). *Denver, Colorado*.
October 2019	Society for Neuroscience Annual Meeting. *Chicago, Illinois*.
October 2018	Janelia Mechanistic Cognitive Neuroscience Workshop. *Ashburn, Virginia*.

A W A R D S

April 2022	International Society for Serotonin Research (ISSR) Biennial Meeting Travel Award